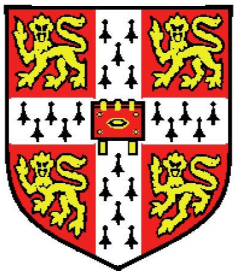
View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Apollo

The Freedom of What We Care About:

Revisiting Frankfurt's Hierarchical Theory of Free Will



Federica Della Grotta

Faculty of Philosophy
King's College
University of Cambridge

This dissertation is submitted for the degree of

Master of Letters

June 2017

For Luca

Abstract

This dissertation concerns the problem of free will. Particularly, it aims to shed light on the hierarchical theory of freedom, firstly presented by Harry Frankfurt in 1971.

The preliminary hypothesis of this dissertation is that the problem of free will is appropriately understood in the terms suggested by the concept of *grounding*. I articulate the relevant claims about freedom in such terms, and I ask consequently what circumstance or condition can make the case that the agent is exercising freedom.

Following this hypothesis, the goal of this work is to show that a revised hierarchical theory of free will constitutes a promising answer to the grounding question about freedom. I develop my arguments in support of this claim in two steps.

First, I propose a critique of Frankfurt's original theory of free will and its later developments. The objective of my analysis is to show that Frankfurt's hierarchical theory of free will is ambiguous between different definitions of freedom and, hence, between different criteria for grounding freedom. At the same time, Frankfurt's later proposals are not successful in the task of justifying free will. As a result, I argue that Frankfurt's theory of freedom lacks a grounding element motivating the ability to exercise free will.

Secondly, I develop an original hierarchical account of free will. Building upon my critique of Frankfurt, my proposal combines the main elements of Frankfurt's original theory (*hierarchical levels of desires, identification*) with new elements from Frankfurt's late reflection about the psychological structure of agents (*caring, volitional essence*). The main aim of my proposal is to equip a hierarchical theory of free will with a new grounding element, the *volitional identity* of the agents. In light of this, I conclude that free will is adequately defined as the agent's wholehearted identification with psychic elements which belong to her volitional identity.

Contents

Introduction **9**

1 Background 9

2 Objectives 16

3 Thesis Structure 17

1 Three concepts of free will **19**

1.1 Outline of the chapter 19

1.2 Frankfurt's hierarchical theory of free will 20

1.3 A ground for freedom 25

1.4 Three concepts of free will 28

1.4.1 Free will as counterfactual causation 31

1.4.2 Free will as actual causation 33

1.4.3 Free will as mere conformity 34

1.5 The freedom of the willing addict 37

1.6 Freedom as a twofold ability 39

1.7 Watson's criticism revisited 42

1.8 Summary of the chapter 48

2 Freedom and Identification **50**

2.1 Outline of the chapter 50

2.2 Identification and Externality 51

2.3 Identification as Decision 56

2.4 Active Identification 61

2.5 Identification as Satisfaction 65

2.6 Wholehearted Identification 67

2.7 Summary of the chapter 72

3 The necessity of freedom **74**

3.1 Outline of the chapter 74

3.2 A revised hierarchical theory of free will 75

3.3 Free will and satisfaction: an ostensible contradiction 81
3.4 Identification, caring and volitional identity 83
3.5 Freedom as satisfaction with one's volitional identity 89
3.6 The problem of ambivalence 94
3.7 A role for practical reason 97
3.8 Manipulated and addicted agents 101
3.9 Summary of the chapter 105

4 Conclusions **107**

Bibliography **110**

Introduction

1 Background

This dissertation concerns the problem of free will. As shown by a huge literature, in philosophy 'free will' is said in many ways. Depending on its definition, free will can require different conditions and explain various and diverse intuitions. The main goal of my dissertation is to shed light on one particular theory of freedom, the hierarchical theory of free will firstly presented in 1971 by Harry Frankfurt. The reason to focus on Frankfurt's account is that I regard it as the most promising account to explain what it means to exercise freedom of the will.

One way to characterize free will is as a kind of 'ability' that a person possesses and that allows her to choose a particular course of action. In light of this characterisation, the problem of free will consists in asking whether we as humans do or do not possess such an ability, and how this ability can be justified. But what does it mean to choose a course of action? A traditional way of understanding this expression is simply the idea that an agent chooses a course of action if such an action is the realisation of a volition of hers. In other words, in order to choose a course of action, an agent has to have the ability of freely doing as she wills (Hobbes 1999 [1654], Locke 1974 [1690], Hume 2007 [1739-40]). However, such a position seems not to distinguish accurately between free action and free will and, in the end, it seems to fail in securing either.

On the one hand, the choice of an agent can be free without guaranteeing that the agent performs a free action. Suppose, for example, that I decide to get

out of my house this morning. Unbeknownst to me, however, someone has locked my front door so that it is impossible for me to get out. In this case, the freedom of my choice does not assure that I can carry out a free action as well. On the other hand, the fact the agent's actions are produced by her desires and volitions seems in itself not enough to guarantee freedom of the will. An agent's volitions can be the result of brainwashing, compulsion, or addiction. In these and similar cases, the action performed by the agent is intuitively not free because it does not stem from a free will.

What is required, then, for the agent to have the ability to choose to act in a free and autonomous way? In this dissertation, I will attempt to reply to this question by understanding it in the terms suggested by the concept of *grounding*. In recent years, the field of philosophy has witnessed an extensive debate about the metaphysical relation of grounding.[1] In this thesis, however, I will not dwell into this technical debate. For the purposes of understanding the problem of free will, I consider the relation of grounding in a broader sense. In particular, I understand grounding in the sense suggested by Sartorio:

> "grounding" [is a] placeholder for some relation of non-causal, ontological dependence between facts in light of which it makes sense to say that certain facts are more basic or more fundamental than others, in that those other facts obtain *because of*, or *in virtue of*, them.
>
> (Sartorio 2016a, 8)

To ask what grounds freedom of the will, then, means to ask what circumstance or condition *can make the case* that the agent is exercising freedom of the will. In other words, when a person is said to choose to act freely, what is it that justifies and secures such an ability?

The aim of this shift of perspective is to contend that when we deal with the problem of free will we are not only looking for necessary and sufficient conditions to justify it. What we are mainly interested in are elements *in virtue of which* a choice or an action can be said free. In this light, the main question about freedom can be described as a question about what is the condition in virtue of which an agent's choice is a free one.[2]

1. For a general overview of the debate about grounding see, for example, Raven (2015).
2. This way of understanding the free will problem has been recently proposed by Sartorio (2016a). See also Bratman (1996) for a similar characterisation of the problem of freedom.

One traditional reply to the question about grounding is that the freedom of the agent is grounded in the presence of *alterative possibilities* available to her. Indeed, a natural way to regard an agent as free at a moment in time is in terms of her capacity to choose between different courses of action at that time. What such an ability assures the agent is a sort of *control* over her decisions: if an agent decides between two opposite alternatives, her control about her decision stems from the fact that she was able, at the time of the decision, to choose the other alternative instead. Indeed, if there is only one course available to her, this means that she is only able to decide one thing, her decision counting as compelled instead of free. Such a traditional view of freedom is endorsed by many incompatibilists about freedom, but also by some compatibilists.[3] According to such a view, then, the ground of freedom lies in the ability of the agent to choose between different alternatives: it is in virtue of such an ability that an agent effectively exercises freedom of the will.

A refusal of the alternative possibilities model was provided by Harry Frankfurt with his 1969 paper "Alternate possibilities and moral responsibility". In this paper, Frankfurt introduced a series of mental scenarios to show that agents are capable of acting freely and being responsible for what they do even if they do not possess different alternatives at the time of the action. The following is a typical Frankfurt-case:

> *Frankfurt scenario*. Ann is deliberating about whether to make a certain choice. Unbeknownst to her, however, a neuroscientist has been monitoring her brain activity. The neuroscientist is now able to predict what choice Ann is about to make, and he is also able to manipulate Ann's brain in a way that assures that the outcome of Ann's deliberation will be B. If Ann will not make the choice to B on her own, the neuroscientist will intervene and force Ann to B. However, Ann makes the choice to B on her own, and the neuroscientist does not intervene in Ann's process of deliberation.

3. The alternative-possibilities model is traditionally defended by incompatibilists such as van Inwagen (1983) and Ginet (1990). For classical compatibilist defenses, see for example Ayer (1954), Davidson (1973), Lehrer (1976), and Lewis (1981). A more recent attempt of defending the alternative possibility model in a compatibilist frame is the so called "new dispositionalism" (Fara 2008, Smith 2003, and Vihvelin 2004 and 2013).

In this scenario, Ann's choice to B is intuitively made on her own, and, as such, Ann is morally responsible for that choice. Crucially, however, at the time of the decision Ann did not had the ability to choose to C, instead. Had Ann chosen to C, the evil neuroscientist would had intervened to correct her process of deliberation and make her choose to B.

Why, in this scenario, one is led to think that Ann is morally responsible for her choice? The reason seems to be that Ann's process of decision was not caused by an external source, e.g. the evil neuroscientist, but by her own desires and motivations. In this respect, in the above scenario Ann is in control of her own decision to B. In other words, in Frankfurt-cases the agent is free and responsible because her choice is produced by the 'right kind' of causes, and as such she intuitively exercises a form of control on her choice, even if she does not have, in practice, alternative possibilities available to her at the moment of the decision.

I would like to suggest to now rephrase the analysis of the above Frankfurt-scenario in terms of the question of grounding. On this approach, the following questions arise: what is it, in the above situation, that makes the case that Ann is acting freely? In virtue of what circumstance(s) is she morally responsible for her choice? Ann does not effectively have the ability to choose otherwise, because the intervention of the evil neuroscientist prevents her from making a different decision. For this reason, the ability to do otherwise cannot count as a proper ground for freedom. What seems to ground freedom in this case is something different.

In particular, Ann's decision to B counts as free because of *how she actually came to make her choice*. Since Ann's choice was motivated by her own desires and reasons, it is the actual sequence of events which leads to Ann's decision which can be said to ground freedom and moral responsibility.[4] In this respect, the presence of the neuroscientist is irrelevant to Ann's freedom and responsibility: since he never interferes with Ann's decision, he is not

[4] Such a position is shared by the advocates of 'Actual-sequence accounts' of freedom, such as Fisher (1994), Fisher and Ravizza (1998), Sartorio (2007, 2012, 2016a, 2016b). The basic claim of these accounts is that the only thing that matters to freedom is the actual sources of the agent's acts, while the presence of alternative in front of the agent is irrelevant to motivate her freedom. Such an insight was suggested by the analysis of Frankfurt's cases, as discussed above. For a detailed description of different varieties of Actual-sequence account see Sartorio (2016a), chs. 1 and 2 and McKenna (2002).

effectively part of the actual process of decision. This seems to explain why Ann is in control of her own decision to B, even if she wasn't able to make a different choice.

One thing to note about Frankfurt-scenarios as the one presented above is that they usually mostly refer to what is required to justify moral responsibility and not directly to freedom of the will. As a matter of fact, a common picture of responsibility assumes that in order to be responsible, an agent needs to satisfy a 'freedom condition' and 'epistemic condition'. According to the former, to be responsible for something, the agent must enjoy some sort of freedom of will or action. According to the latter, to be responsible for something one needs also to be aware of what one does, and of the moral implications of what one does.

Since Frankfurt-scenarios mostly refer to freedom in terms of its relevance for moral responsibility - hence addressing the question of what moral responsibility is grounded in- Frankfurt's answer as to what is needed to ground freedom of the will itself seems to be less clear.

Frankfurt presented his account of freedom of the will in his 1971 paper "Freedom of the will and the concept of a person". Here, he understands the ability to exercise a free will in strict connection with the complex psychological structure of agents. On Frankfurt's view, being a human person is essentially linked to the ability to have desires and motives of a second order. To have a second-order desire means that one has the ability to want to have or not to have certain desires.

For Frankfurt, having free will is strictly linked to the possession of such an ability. Frankfurt claims that when one acts in accordance with a desire that she wants to have, she is acting with a free will. On the contrary, when one acts on desires which she does not recognize at a higher level of reflection as 'her desires', she is not acting with free will. In his 1971 essay, Frankfurt firstly considers cases in which an agent is driven by motives which she recognizes as alien to her motivational structure. The analysis of such cases suggests that agents involved in these scenarios - such as, for example, addicts or compulsive agents - do not recognize their desires as free or autonomous. In contrast, Frankfurt argues, it seems that to be a free and autonomous agent one has to

be directed by motives and desires that she approves and that she wants to be part of her motivational structure.

Accordingly, the core claim of Frankfurt's account of freedom is that free will is strictly dependent on how the agent's desires are organized when they motivate a choice. As with the case of moral responsibility considered in Frankfurt (1969), it seems that what counts for freedom of the will is not the presence of alternative possibilities available to the agent in the moment of the choice. Rather, the will of the agent is free if it is formed by effective desires which are brought about 'in the right way' in the motivational organization of the agent.

What is, however, such a 'right way' and why is it significant for freedom of the will? The answer to this question is crucial in order to corroborate Frankfurt's theory of freedom. One way to tentatively reply to this question is by rephrasing it again in the terms offered by the notion of grounding. In this sense, one can ask in which way the organization of the agent's desires can ground freedom of the will. In other words, how can the motivational structure of the agent make it the case that the agent is exercising freedom of the will?

It seems that for Frankfurt (1971), the right answer to such a question is that free will is principally grounded in the hierarchical organization of the will and, in particular, in the relation of dependence between lower-order and higher-order desires. More specifically, for the will to be free, the lower-order efficacious desires of a person have to be in accord with, or in some way controlled by, higher-order desires. In light of this, Frankfurt's first answer is that free will is grounded in the role of higher-order volitions in the agent's will. Particularly, Frankfurt's answer at the time of his 1971 paper is that some effective desires are brought about 'in the right way' because they are brought about by the higher-order volitions that the agent endorses. As a consequence, an agent exercises a free will when she is driven by desires which are endorsed by herself at a higher level of reflection.

However, Frankfurt's (1971) arguments to justify these claims are not always univocal. One problem with Frankfurt's theory is that he is not clear about the precise role which is played in this picture by higher-order desires. On the one hand, it is not clear whether the higher-order desires must only be in accord with or also be a direct cause of the lower order desires, or maybe both (Shatz 1985, Sartorio 2016, Beebe 2013). On the other hand, it is not clear if

the role that Frankfurt assigns to higher-level desires is successful in this task. Indeed, a traditional objection to Frankfurt's theory is that higher-order desires in themselves are nothing special, and the idea that their role in the agent's motivation can ground freedom of the will is simply misleading (Watson 1987, Bratman 1996).

These and other objections led Frankfurt to subsequently reject the view that free desires are simply reducible to endorsing higher-order volitions of the agent. Much of Frankfurt's work since 1971 can be interpreted as an attempt to provide a different answer to the question of what grounds freedom of the will. Starting from Frankfurt (1977), he replies to this question by focusing on the notion of *identification*. Roughly, identification is a psychic process by which the agent's motivational elements, in particular her desires, become in some sense 'her own'. Frankfurt now suggests that a desire is free when an agent identifies with it. According to such a view, when the agent is moved by desires with which she 'identifies', her desires are not external to her motivational structure, and thus she is exercising a free will. On the contrary, when the desire that moves the agent is perceived by the agent as 'alien', as one with which she does not identify (as, for example, in case of addictions and compulsions), she is not driven by a will that is free. This latter notion of identification, however, is also threatened by a number of objections. The development of his theory from his 1977 to his 1992 paper thus aims at revising such a notion in different directions in order to meet these objections.[5]

For these reasons, Frankfurt's theory is not clear about what is the element which can ground freedom of the will in his account (Bratman 1996 and 2002; Watson 1975 and 1987; Sartorio 2016a). Throughout a number of works, Frankfurt revised his account of freedom and provided different answers to this question (*higher-order volitions, identification, satisfaction*). The presence of different answers in Frankfurt's account of freedom seems to be a source of confusion, when not ambiguity, in Frankfurt's theory.

In recent years, however, Frankfurt's reflections have followed a new direction as regards to freedom and autonomy. In parallel with the development of notions such as caring, love and volitional identity within his

[5] See particularly chapter 2, section 2.2 for a discussion of the classical objections to Frankfurt's notion of *identification*.

philosophy of action, Frankfurt now claims that "[a] person acts autonomously only when his volitions derive from the essential character of his will" (Frankfurt 1982, 132). Such a reflection suggests a different understanding of the notion of freedom, connecting it to the notion of volitional essence. The agent's desires are then free when they are 'essential' to the subject's volitional identity. The fascinating claim made by Frankfurt is that some of the agent's desires form, in some way, what can be called her volitional essence. Such a volitional essence is so powerful that it impedes the agent from identifying with desires which are not part of her volitional identity. For the agent to identify with some desires and not with others is, indeed, a kind of necessity, a volitional necessity. The interesting question Frankfurt raises concerns the way in which the idea of a volitional essence is connected with freedom and autonomy.

One of the aims of my dissertation is to provide an answer to this question, that is, to motivate the claim that freedom is in some way dependent on the agents' volitional essence. In order to do so, I propose a modified hierarchical theory of free will, which revises elements from Frankfurt's account presented in 1971 in light of Frankfurt's recent works. My intention is to build from Frankfurt's rich description of the psychology of agency in order to derive a hierarchical theory of free will. In particular, my account of freedom justifies the introduction of volitional identity as a new ground for freedom. Free will, in this sense, is understood as the agent's wholehearted identification with psychic elements which belong to her volitional identity.

2 Objectives

The main objective of this dissertation is to argue that a revised hierarchical theory of free will is a promising answer to the grounding question about freedom. I will develop my arguments to support this claim in two steps:

A. First, I propose an analysis of Frankfurt's original theory of free will and its later developments. The objective of my analysis is to show that Frankfurt's original hierarchical theory of free will is ambiguous between different definitions of freedom, and that it does not have the resources to address some very well-known objections against it. At the

same time, I argue that the later developments of Frankfurt's account of freedom, that are intended as providing a reply to the objections and a ground to the theory, are not successful in that task. As a result, I argue that Frankfurt's concept of freedom is left without a secure grounding element to explain the ability to exercise free will.

B. Secondly, my objective is to develop a revised hierarchical theory of free will. Building upon my analysis of Frankfurt's account of freedom, my account combines the main elements of Frankfurt's original theory (*hierarchical levels of desires*, *identification*) with new elements from Frankfurt's late reflection about the psychological structure of agents (*caring*, *volitional essence*). The main aim of my proposal is to equip a hierarchical theory of free will with a different and new grounding element, the *volitional identity* of the agents.

3 Thesis Structure

This dissertation is structured as follows. In chapter 1 (*Three concepts of free will*), I discuss Frankfurt's (1971) theory of free will. I argue that Frankfurt's classical theory of free will grounds freedom not just by appealing to the endorsement of higher-order volitions, but by tightening the conditions of control of the agent on her desires, so as to ensure that the agent *directly controls* her motivational structure. Such a move aims to show that if a desire is internally controlled by the agent's volitions, then her desires are brought about 'in the right way' in the motivational organisation of the agent, thus counting as free in Frankfurt's sense of the term. However, I suggest, the condition of control that Frankfurt (1971) provides is not univocal, and gives rise to three different conceptions of free will. I show that also higher-order volitions, even if enhanced with a distinction between three concepts of free will, do not work as a grounding for freedom because they are vulnerable to an influential objection originally made by Gary Watson (Watson 1975).

In chapter 2 (*Freedom and Identification*), I discuss Frankfurt's claim that identification is sufficient to secure freedom. I argue that Frankfurt presents two different notions of identification as a way to ground his account of freedom. I call them *Active Identification* and *Wholehearted Identification*. In a first

time, he argues for a notion of active identification with the introduction of the concept of decision. In a second time, he rejects this account and argues for a notion of wholehearted identification with the introduction of the concept of satisfaction. I argue that both such approaches are doomed to failure and that, as such, Frankfurt's notion of identification is unable to act as a ground for a hierarchical theory of freedom.

In chapter 3 (*The necessity of freedom*), I present a revised hierarchical account of free will. The main claim of my proposal is that freedom is dependent on the agent's volitional identity, which acts as a proper ground for freedom of the will. First, I motivate the agent's identification with her desires by proposing a different notion of identification, which make use of the notion of satisfaction, but it is grounded in the concepts of volitional identity and caring. I propose to make use of such a notion of identification to ground a hierarchical account of free will. Free will, in this sense, is described as the agent's wholehearted identification with psychic elements which belong to her volitional identity.

Chapter 4 concludes.

1

Three concepts of free will

1.1 Outline of the chapter

In this chapter, I discuss Frankfurt's (1971) theory of free will. I argue that Frankfurt's classical theory of free will grounds freedom not just by appealing to the endorsement of higher-order volitions, but by tightening the conditions of control that the agent has over her desires, so as to ensure that the agent *directly controls* her motivational structure Such a move aims to show that if a desire is internally controlled by the agent's volitions, then her desires are brought about 'in the right way' in the motivational organisation of the agent, counting as free in Frankfurt's sense of the term (section 1.3). However, I suggest that the condition of control that Frankfurt (1971) provides is not univocal, but rather gives rise to *three different definitions of free will* and, hence, to different criteria for grounding freedom (section 1.4). Distinguishing between these three concepts of freedom serves a number of additional purposes in this chapter. Specifically, I offer an original formulation of the problem of the willing addict (section 1.5). Moreover, I make use of this distinction to revisit Watson's (1975) classical objection to Frankfurt's theory (section 1.7). I conclude that higher-order volitions, even if enhanced with a distinction between three concepts of free will, do not work as a grounding for freedom because they remain vulnerable to Watson's objection (section 1.7).

1.2 Frankfurt's hierarchical theory of free will

Frankfurt's (1971) paper "Freedom of the Will and the Concept of a Person" presents an original account of free will. This account is tightly related to Frankfurt's own conception of the psychological structure of a person's motivation. In this paper, he also introduces the *hierarchical theory of desires,* which is at the core of his account of human psychology and action. Frankfurt (1971) argues that having free will and being a person are significantly tied together: in particular, he suggests that free will is the outcome of the specific hierarchical organisation of a person's will.

The main elements which constitute the will of a person are *desires*. Frankfurt thinks that human beings have a certain motivational complexity, and as such their will is formed by desires of different levels. On this account, persons possess a primary level of motivation formed by first-order desires. First-order desires are desires which have as their objects an action. Frankfurt contends that first-order desires can be expressed in statements of the form:

[P] wants to X

where 'to X' refers to an action (1971, 13).[6] If, for example, I want to go for a run, this is a first-order desire of mine. In turn, among first-order desires, Frankfurt distinguishes between desires which are effective and those which are not. Effective desires are those that are able to move a person, as Frankfurt puts it, "all the way to action" (1971, 14). Effective first-order desires primarily constitute the *will* of an agent[7]. This, however, is a not an exclusive characteristic of human persons. On the contrary, other creatures such as many non-human

[6] Page numbers are from the reprints of the articles in Frankfurt (1988) and (1999), despite the fact that the papers are cited in the original date.

[7] Note that Frankfurt uses the verbs 'want' and 'desire' as synonyms. Indeed, Frankfurt (1971) suggests the following usage: "A wants to X may mean to convey that it is his desire that is motivating or moving A to do what he is actually doing or that A will in fact be moved by this desire (unless he changes his mind) when he acts" (14). One possible objection to this choice is that it does not take into account the difference between mere desires and evaluations. It could make sense to say that, for example, though I desire to eat junk food, I do not actually *want* it because I want to live healthily. In this case, the verb 'want' seems to be directly related to an evaluation made by the person. Because of this, it seems that one can actually fail to want what one desires and the other way around (see Watson 1975 for a similar point).

animals are able to conduct actions by following effective desires. But if this is true, what, if anything, does characterise the organization of a *person's* will?

According to Frankfurt, what characterises the structure of a person's will is the circumstance that she also possesses second-order desires. Such second-order desires are said to have first-order desires as their objects (1971, 14). For example, I can have the desire to desire to get fit and exercise daily, and this can result in the desire to have the desire to go for a run. In this case, I want to have a first-order desire and, moreover, I want such a desire to be effective, so as to move myself all the way down to the action of going out for a run.

The objects of second-order desires are first-order desires, and I can want them to be effective or not: I can simply have a second-order desire for having a certain kind of desire without wanting to act on it. For this reason, Frankfurt recognises a further distinction among second-order desires. It is possible for me to have a second-order desire for a first-order desire, and still do not want such a desire to be part of my will. Frankfurt's example is very revelatory. He imagines a psychotherapist who desires to have a first-order desire to take a drug, in order to understand better her patients' feelings. However, she is very resolute in wanting to avoid really taking the drug: what she wants is only to feel the desire, without being moved to act by such a desire, and without considering such a desire as a part of her will (1971, 9). In more normal cases, agents have second-order desires which they want to be effective, which they want to be part of their will. As in the previous example, if I have a desire to get fit and to go out for a run, in this case I want my desire for a healthy body to be completely effective: I want to have a will which includes the desire to have a healthy body. Such kind of second-order desires are called by Frankfurt *second-order volitions.* As he explains:

> Someone has a desire of the second order either when he wants simply to have as certain desire or when he wants a certain desire to be his will. In situations of the latter kind, I shall call his second-order desires "second-order volitions" or "volitions of the second order". Now it is having second-order volitions, and not having second-order desires generally, that I regard as essential to being a person.
>
> (Frankfurt 1971, 16)

According to Frankfurt, only persons possess second-order volitions. The distinguishing feature of a person's will, then, is not just the possibility of having second-order desires but that of having second-order volitions as well.

The presence of second-order volitions in a person's motivation is fundamental. Indeed, only persons can have the desire to be moved by desires which are *different* from the ones which actually move them. That is, they can want a certain desire to become their will. For this reason, the uniqueness of persons can be then traced back to their natural capacity to *reflect on themselves* and on their desires and wants: this means that they are at the same time reflective and the object of this reflection (cf. Buss and Overton 2002). Because of this capacity, it is possible for persons to find a discrepancy between these two dimensions. In other words, it is possible that they find that they have desires which they would prefer not to have or manifest, because they would like to possess a different will from the one they discover in themselves by the process of self-reflection. Such an ability, Frankfurt claims, is the essence of being a person, and an individual without it is a 'wanton'. What is missing in a wanton is not, Frankfurt argues, a sort of rationality, but instead the exact ability to reflect on her motivational structure and accordingly form higher-order volitions (Frankfurt 1971, 16).

Frankfurt's characterisation of the wanton indicates another crucial aspect of his theory. Namely, that which makes persons 'persons' lies primarily in the structure of the will. What is able to move persons to action is the hierarchical structure of their will rather than their rational abilities.[8] In Frankfurt's picture, the reasoning ability only has an instrumental value.[9] At the same time, being a person presupposes rationality: indeed, rational abilities allow a person to be critically aware of her own motivational structure by manifesting second-order volitions. Accordingly, the structure of a person's will presupposes that she is a rational being (Frankfurt 1971, 17).

Building upon these distinctions, Frankfurt (1971) develops a hierarchical theory of freedom. According to Frankfurt, there is a close connection between

[8] Frankfurt's conception of reason is, in this respect, in line with the Humean claim that pure reason does not move us, but is instead "the slave of passions", a merely instrumental ability. "Reason alone can never be a motive to any action of the will," and reason alone "can never oppose passion in the direction of the will" (Hume 1748 [2007], 413).

[9] Frankfurt motivates his view on practical reason in his 2004 book *The Reasons of Love*. I discuss his approach in more detail in chapter 3, section 3.7.

the structure of a person's will and the possession of a distinguishing feature of the human condition: the ability to exercise freedom of the will.

In Frankfurt's words:

> It is only because a person has volitions of the second order that he is capable both of enjoying and of lacking freedom of the will. The concept of a person is not only, then, the concept of a type of entity that has both first-order desires and volitions of the second order. It can also be construed as the concept of a type of entity for whom the freedom of its will may be a problem.
>
> (Frankfurt 1971, 19)

In this passage, Frankfurt makes two connected remarks. First, being capable of having second-order volitions and of being reflective about her will is a pre-condition for a person's ability to enjoy or to lack free will. Secondly, and consequently, it is possible to identify persons not only as those entities that possess second-order volitions, but also as the type of entities for whom enjoying or lacking freedom of the will may be a problem. Indeed, as they lack the ability of self-reflection, other kinds of entities, such as non-human animals, wantons, and arguably even very young children, do not possess the capacity of exercising or not exercising free will.

To have free will is then, in Frankfurt's terminology, a capacity which involves desires of the second-order, and someone has second-order volitions when he wants "a certain desire to be his will". As compared to the classical compatibilist approach, the advantage of this definition of free will is, as Frankfurt suggests, that classical compatibilism is only able to account for freedom of action but not for freedom of the will. To say that to be free is to be able *to do what one wants* seems to be sufficient condition for freedom of action: if one is able to perform, without impediments, a willed action, she is effectively doing what she wants, thus enjoying freedom of action. In this case, if nothing impedes the relation between the will of an agent and the action she performs, then the agent is exercising freedom of action.

On the other hand, to have free will is, in Frankfurt's terminology, a capacity linked to the higher-level of the self ("The question of the freedom of his will does not concern the relation between what he does and what he wants

to do. Rather, it concerns his desires themselves", Frankfurt 1971, 20). Frankfurt proposes to understand free will in close analogy with the understanding of freedom of action. In this light, it is possible to claim that freedom of the will is the freedom of wanting, through second-order volitions, what one *wants to want*. An agent's *will* is free, then, when nothing impedes the relation between volition and will (cf. McKenna 2002).

Frankfurt recognises at least two possible complexities for his theory. First, second-order volitions, as they are desires, can conflict among themselves. But in this case, until the conflict is resolved, it would be uncertain whether in acting in accord or in conflict with any particular volition an agent is acting of her own free will. Secondly, according to Frankfurt, it is possible to imagine that a person may have, especially in the case of a conflict between her second-order volitions, volitions of a still higher order than the second ones (Frankfurt 1971, 21). The question then becomes what, in such a case, can prevent a person from obstinately refusing to identify herself with one of her second-order volitions, so as to act with a will that she wants to have. To answer this question, Frankfurt introduces the notion of *identification*, one that he will significantly develop in his later writings*:*

> It is possible, however, to terminate such a series of acts without cutting it off arbitrarily. When a person identifies himself decisively with one of his first-order desires, this commitment "resounds" throughout the potentially endless array of higher orders.
>
> (Frankfurt 1971, 21)

What Frankfurt further requires for justifying freedom of the will is not only the fact that the agent satisfies the hierarchical mesh between volition and will, but also that she *decisively identifies* herself with one of her first-order desires. In so doing, her will becomes constituted by the volitions aligned with her identification, that is, with what at a higher-order of reflection she wants her will to be. To enjoy free will, then, is a matter of "satisfaction" of those desires with which a person decisively identifies herself, while a lack of such identification would entail the absence of a free will.

In light of the above treatment, Frankfurt (1971) proposes the following characterisation of freedom of the will:

Freedom of the will: a person possesses freedom of the will if she is free to want what she wants to want.

On this formulation, 'wants' refers to higher-order desires whose objects are the first-order desires with which the agent identifies.

1.3 A ground for freedom

In the Introduction, I argued for the opportunity to understand the problem of free will in terms suggested by the notion of *grounding*.[10] As proposed by Sartorio (2016a), the relation of grounding holds that "it makes sense to say that certain facts are more basic… than others, in that those other facts obtain *because of*, or *in virtue of*, them" (8). In light of this, to ask what grounds freedom of the will means to ask what circumstance or condition can make the case that the agent is exercising freedom of the will. When a person is said to choose to act freely, what is it that justifies and secures such an ability? In this section, I ask how hierarchical theories of free will, and particularly Frankfurt's account, are able to reply to this question. I will do so by comparing Frankfurt's theory with classical compatibilist accounts of freedom.

Classical compatibilist theories of freedom hold that a person acts freely if there are no external impairments to her doing what she wants to do (e.g., Hobbes 1999 [1654], Locke 1974 [1690], Hume 2007 [1739-40]). For example, if a person is in chains or locked in a room, she cannot do what she most wants to do, thus she cannot act freely. In contrast, if a person is able to simply act as she desires, she is acting freely. According to classical compatibilism such an analysis is sufficient to also explain freedom of the will, because 'freedom' or 'liberty' can be equally applied to actions or choices: when an agent has no impediments to her doing what she wants to do, this implies that her choices are as well free, and that she is exercising freedom of the will.[11] Such a 'thin'

[10] See Introduction, section 1, 2-3.

[11] See for example Hume's definition of freedom, according to which to be free is simply to perform actions in line with desires: "[W]hat is meant by liberty when applied to voluntary actions? We cannot surely mean that actions have so little connection with motives, inclinations, and circumstances that one does not follow with a certain degree of uniformity from the other, and that one affords no inference by which we can conclude the existence of

account of freedom is straightforwardly compatible with determinism, and it additionally has the merit of describing freedom of the will by just considering ordinary aspects of human agency.

In more rigorous terms, classical compatibilism requires for freedom of the will the fulfilment of two conditions. The first condition is supposed to guarantee the self-determination of the agent, and says that if the agent's behaviour is produced by the agent's wants, beliefs, and desires, then she is acting with a free will. The second condition provides the agent with alternative possibilities, and says that if the agent had been motivated differently, her behaviour would have been different (cf. Levin 1979, 238).

In light of this, classical compatibilism grounds freedom in two different elements. On the one hand, freedom is justified by the link between the actual motivation and the actions of the agent: an agent acts freely if her behaviour is directly caused by her desires and wants, without any external impediment. On the other hand, freedom for classical compatibilists is grounded in the presence of counterfactual alternative possibilities for the agent or, in other words, in the connection between the agent's counterfactual motivation and her acts: had the agent been motivated differently, she would have acted in a different way.

Notoriously, such a position presents immediate problems. That the agent's behaviour is produced by her desires and wants does not seem to be enough to guarantee freedom, because the agent's desires themselves can be the origin of alienation. In cases of compulsion or addiction, for example, what prevents the agent from exercising freedom of the will is exactly the presence, in her motivational structure, of 'wrong' desires. Indeed, when an addict takes a drug, she could do so in accordance with and because of her desire for the drug. The addict could also possess alternative possibilities: had she wanted to act otherwise, she would have done so. If both of these conditions are respected, according to the classical compatibilist such an addict is effectively acting freely. It is easy to see that such a conclusion is strongly

the other. For these are plain and acknowledged matters of fact. By liberty, then, we can only mean *a power of acting or not acting according to the determinations of the will;* that is, if we choose to remain at rest, we may; if we choose to move, we also may. Now this hypothetical liberty is universally allowed to belong to everyone who is not a prisoner and in chains. Here then is no subject of dispute" (Hume 1748 [2007], VIII, Part I, 104).

counterintuitive.[12] It seems, then, that the grounding elements provided by compatibilism are not sufficient to secure freedom of the will.

By contrast, Frankfurt's hierarchical theory seems to be better equipped to deal with these problems. The general strategy adopted by hierarchical theories of free will is one of constructing a model of the agent's motivational organisation as composed of different levels of desires.[13] In this sense, Frankfurt shows that alien desires or intentions (e.g. the desires of an addicted agent) are the result of an incorrect mesh between different levels of the person's motivational structure. The analysis of addicts or compulsive agents shows that they do not recognise their desires as free or autonomous. In contrast, Frankfurt claims, it seems that to be free and autonomous an agent has to be directed by motives and desires that she approves and that she wants to be part of her motivational structure. Accordingly, for Frankfurt free will is strictly dependent on *how* the agent's desires are organised when they motivate a choice, and not simply on the fact that such desires and wants cause the agent's actions. In this light, the will of the agent is free if it is formed by effective desires which are brought about 'in the right way' in the motivational organisation of the agent.

In which way, then, does the organisation of the agent's desires motivate freedom of the will in a way which is different from the one offered by classical compatibilism? Shatz (1985) suggests that Frankfurt's overall strategy to deal with the classic objections to compatibilism is to tighten the conditions of free will so as to "ensure that the agent controls his motivational structure in precisely the way that his first-level motivational structure controlled his behaviour in the original, unrefined account" (459). More specifically, for the will to be free, the lower-order efficacious desires of a person have to be in accord, or be in some way controlled by higher-order desires. In this light,

[12] The respect of the second condition is in itself strongly problematic in a compatibilist scenario, as the vast literature on the subject shows. In this dissertation, I will not discuss this condition in detail. Instead, I will limit my arguments to the first condition. On a general note, however, it seems that to analyse "I would have done otherwise if I wanted to" in terms of choices and desires leads to a regress, because it is always possible to ask the question "But could the person have chosen differently?" (cf. Broad 1934, Taylor 1992). In this sense, such an analysis seems to lead to an infinite regress of ever higher-order counterfactuals (see Davidson 1980, 69).

[13] Among hierarchical theories of free will, the most renowned are probably Dworkin (1970); Frankfurt (1971), (1975), (1987); Neely (1974); Levin (1979); Lehrer (1980); Watson (1975), (1977); and Zimmerman (1981).

Frankfurt's answer is that free will is secured by the role of higher-order volitions in the agent's will.

In what follows, I argue that such a move allows Frankfurt to define a different grounding element for freedom. Indeed, for Frankfurt (1971), free will is mostly grounded in the relation of dependence between lower-order and higher-order desires. Frankfurt defines a different ground for freedom by tightening the conditions of control of the agent on her desires so as to ensure that the agent *directly controls* her motivational structure. Such a move aims to show that if a desire is internally controlled by the agent's volitions, then her desires are brought about 'in the right way' in the motivational organisation of the agent, counting as free in Frankfurt's sense of the term. However, I suggest that the condition of control that Frankfurt (1971) provides is not univocal, but rather gives rise to *three different definitions of free will* and, hence, to different criteria for grounding freedom.

Distinguishing between these three concepts of freedom will serve a number of additional purposes later in this chapter. Specifically, such a distinction will allow me to offer an original formulation of the problem of the willing addict (section 1.5).[14] Moreover, I will make use of this distinction to revisit Watson's (1975) classical objection to Frankfurt's theory (section 1.7).

1.4 Three concepts of free will

In this section, I argue that in Frankfurt (1971) the *condition of control* which is supposed to link the agent's higher-order motivational elements with the lower-order ones is not spelled out in a univocal way. I argue that such an ambiguity in the definition of a control condition gives rise to three different concepts of free will. As a result, I suggest that Frankfurt (1971) also presents diverse grounding elements for free will.

Consider the following passage as an illustration of the different concepts of free will I would like to spell out:

[14] In chapter 3, section 3.8, I offer as well an original solution to the problem of the willing's addict freedom. There, I suggest that the willing addict is only allowed to exercise free will in a weaker sense than the one claimed by Frankfurt (1971).

> Just as the question about the freedom of an agent's action has to do with whether it is the action he wants to perform, so the question about the freedom of his will has to do with whether it is the will he wants to have. It is in securing the conformity of his will to his second-order volitions, then, that a person exercises freedom of the will. And it is in the discrepancy between his will and his second-order volitions, *or* in his awareness that their coincidence is not his own doing but only a happy chance, that a person who does not have this freedom feels its lack.
>
> (Frankfurt 1971, 20, *my emphasis*)

In this passage, Frankfurt describes the conditions under which a person does exercise or does not exercise freedom of the will. On the positive case, Frankfurt defines free will as a conformity between the will and the second-order volitions of an agent: if they align, and so the person possesses the will she wants to have, she exercises free will. On the negative case, however, he seems to assume two alternative circumstances. First, if there is a 'discrepancy' between the will and the second-order volitions, the person is not exercising free will. Secondly, and alternatively, if the will and the second-order volitions align, but such alignment is only a 'happy chance' and not the person's own doing, the person does not exercise freedom of the will.

Accordingly, the passage seems to conflate two different concepts of free will:

(a) A person possesses free will if an only if she can make her will and second-order volitions conform.
(b) A person possesses free will if and only if her will and her second-order volitions conform.[15]

[15] Another way to express these two meanings of free will by using Frankfurt's own terminology is as follows:

(a1) A person possesses free will if and only if she is free to want what she wants to want.
(b1) A person possesses free will if and only if she wants what she wants to want.

Condition (a1) suggests that a person exercises free will only if her will is in itself free, that is, if and only if she is free to want something different from the will she already has. Condition (b1), on the other hand, suggests that for a person to exercise free will it is sufficient that she actually desire in conformity with her will. In chapter 3, section 3.2, I will refer to conditions (a1) and (b1) instead of (a) and (b).

The difference between these two concepts of free will can be expressed by noticing that while (a) suggests that there should be a direct relation between higher and lower motivational elements of the agent – i.e. higher level motivational elements should cause or provide a reason for the lower level ones, (b) suggests that second order volitions should be merely 'in accord' with the first-order desires which constitute the will of an agent.

Moreover, I argue that Frankfurt (1971) is also ambiguous between two different readings of concept (a), that is, the condition that a person possesses free will if she can make her will and second-order volitions conform. More precisely, meaning (a) is fulfilled by the respect of two different control conditions:

(i) *Actual control.* The relation of control between motivational elements of the agent has to be satisfied only in an actual scenario but not in a counterfactual one.

(ii) *Counterfactual control.* The relation of control between higher-order desires and lower-level ones has to be satisfied both in an actual and in a counterfactual scenario.

The difference between condition (i) and (ii) can be understood as the difference between having actual causal power only in the actual moment or across a wide range of circumstances. In the former, the agent exercises control on her will by exercising causal control in the actual scenario, while in the latter, such a condition of control has to be respected in different conditional circumstances.

Building on these remarks, my claim is that Frankfurt (1971) proposes three different concepts of free will:

(c1) *Free will as counterfactual causation,* according to which the condition of causal control has to be respected both in an actual and in a counterfactual scenario.

(c2) *Free will as actual causation,* according to which the condition of causal control has to be satisfied only in an actual scenario.

(c3) *Free will as mere conformity*, according to which the condition of control does not have to be satisfied.

In the remainder of this section, I will spell out the differences between these three concepts.[16] I will also suggest that these three concepts of free will provide Frankfurt's theory with different grounding elements for freedom.

1.4.1 Free will as counterfactual causation

As an illustration of (c1), that is, *free will as counterfactual causation*, one can consider the following passage (Frankfurt 1971, 24):

> A person's will is free only if he has the will he wants. This means that with regard to any of his first-order desires, he is free either to make that desire his will or to make some other first-order desire his will instead. Whatever his will, then, the will of the person whose will is free could have been otherwise; he could have done otherwise than constitute the will as he did.

This concept of free will is meant to run in parallel with the classical compatibilist definition of freedom. As discussed in section 1.3, classical compatibilism requires for freedom of the will the fulfilment of two conditions. On the one hand, it requires that the agent's behaviour is produced by her wants, beliefs, desires. On the other hand, it requires the presence for the agent of counterfactual alternative possibilities, and says that if the agent had been motivated differently, her behaviour would have been different.

In the same way, according to Frankfurt a person possesses free will only if she is free to have the will she wants. This definition implies the presence of two different conditions for free will. On the one hand, it implies the presence

[16] The distinction that I am proposing between different conceptions of freedom has been neglected in the traditional literature about Frankfurt's hierarchical theory. Rostbøll (2004) tries to detect the presence of different ideas of freedom in Frankfurt's account, and claims that Frankfurt distinguishes between an *Avoidability Conception* and a *Satisfaction Conception* of freedom. I have developed my arguments independently from Rostbøll, and the distinction I propose is different from the one he articulates, and serve different purposes in my argument.

of alternatives, that is, it implies the possibility that the person is able to choose, in an alternative scenario, between different wills.[17] Freedom in this picture is then constituted as a dependency relation between first-order desires and second-order volitions, and the presence of alternatives is related to the higher-order organisation of a person, i.e. her second-order volitions. Moreover, this definition suggests that a person is free if she is in some way able to *produce* the alignment between her will and her second-order volitions. If this happens and this is not just a happy chance, then the person's will aligns with her second-order volitions and she exercises freedom of the will.

In which way, however, is the person able to produce the alignment in her will? Here it seems that three possibilities stand. First, it can be the case that the lower-order desires of the agent cause the higher-order volitions to align. Secondly, it can be the case that the higher-order volitions cause the lower-order volitions to align. Thirdly, it can be the case that the alignment between higher-order and lower-order volitions is produced by something else, as for example by the agent herself and not, specifically, by her volitions of any order. entertaining

In my interpretation, Frankfurt (1971) is inclined to favour the second scenario. In his account, second-order volitions seem to show a *causal* power: they are able to make some first order desires effective and/or they seem able to bring about different first-order desires. To motivate this, consider the following remarks by Frankfurt (1971, 25, *my emphasis*) regarding the willing addict's situation:

> I am inclined to understand his situation as involving the overdetermination of his first-order desire to take the drug. This desire is his effective desire because he is physiologically addicted. But it is his effective desire also *because he wants it to be*. His will is outside his control, but, *by his second-order desire* that his desire for the drug should be effective, he has made this will his own.

In this passage, Frankfurt explains the alignment in the agent's will by postulating an *overdetermination* of the willing addict's first-order desires to

[17] It would be interesting to ask whether the presence of a causal relation between higher-level desires and lower-level ones by itself requires entertaining a counterfactual scenario. However, the answer to this question is not clear. For this reason, I will consider the condition of *internal control* and the one of *counterfactual control* as two different constraints on a theory of free will.

take the drug: her desire is effective because she is addicted to the drug but, at the same time, it is her effective desire because it is produced by her second-order desire to take the drug.[18] By extending Frankfurt's argument to normal, non-addicted agents, it seems that the agent's effective will is produced, at least partially, by the causal action of second-order volitions.

According to these considerations, it is possible to formulate the first definition of free will proposed by Frankfurt:

> *(Def. 1) Free will as counterfactual causation.* A person exercises free will if she is free to have the will she wants, and she could have constituted her will otherwise than she did.

Summing up, to endorse *(Def. 1)*, it is indeed necessary to endorse the two different conditions of (i) *actual control* and (ii) *counterfactual control*.[19]

My further claim is that these two conditions provide *(Def.1)* with two grounding elements for freedom. On the one hand, according to (i) free will is grounded in the role of higher-order volitions as actual causal elements. On the other hand, following (ii), free will is secured by the role of higher-order volitions as counterfactual possibilities.

1.4.2 Free will as actual causation

In spite of the latter conclusion, in this section I argue that Frankfurt (1971) additionally envisages a slightly different conception of free will, as expressed by condition (c2). To illustrate this conception, that is *free will as actual causation*, consider the following passage:

> Suppose that a person has done what he wanted to do, that he did it *because he wanted to do it*, and that the will by which he was moved when he did it was his will because it was the will he wanted. Then he did it freely and of his own free will. Even supposing that he could have done otherwise, he would not have done otherwise; and even supposing that he could have

[18] At a later time, Frankfurt attempts to explain the alignment in the agent's will by making use of the third possibility, that is, by arguing that the alignment between lower-order and higher-order volitions is produced by something else than the agent's volitions. In particular, he shifts the perspective from the casual efficacy of second-order volitions to an activity performed by the agent with the act of *decision*. I discuss this further possibility in chapter 2, section 2.3, where I show that Frankfurt's proposal does not work to ground free will.

[19] Cf. section 1.4, 20.

> had a different will, he would not have wanted his will to differ from what it was.
>
> (Frankfurt 1971, 24, *my emphasis*)

The agent described in this passage exercises freedom of the will. In this case, however, even if the counterfactual ability could be present, nonetheless this definition of free will *does not depend* on such a counterfactual ability. According to this conception of free will, a person's will is free if her second-order volitions are internally related to her will. Moreover, even if the person could be able to change her will, nonetheless she never has an interest in changing it.

In this light, what Frankfurt requires for exercising free will in this second case is only that (i), i.e. *actual control*, is respected, while it is not necessary that (ii), i.e. *counterfactual control,* is also respected. That is, the agent is exercising freedom of the will because she has an internal control over her will, insofar as her second-order desires are causally related to her lower-level desires. At the same time, she does not have to possess counterfactual control.

A second definition of free will can be formulated accordingly:

> *(Def. 2) Free will as actual causation.* A person exercises free will if she has the will she wants because it is the will she wants to have, and she has no interest in changing it.

As a consequence, *(Def.2)* provides a slightly different grounding element for free will. According to *(Def. 2)*, indeed, freedom is only grounded in the role of higher-order desires as actual causal elements.

1.4.3 Free will as mere conformity

It is possible to detect in Frankfurt (1971) a third concept of free will. According to conception (c3), that is *free will as mere conformity*, a person enjoys free will if her will is in accord with her second order volitions, that is if her will *conforms* to her second-order volitions: if she wants what she wants to want. Conversely, if there is a 'discrepancy' between the person's will and her second order volitions, the person is not exercising freedom of the will.

As an exemplification of conception (c3), consider the following passages:

> The unwilling addict's will is not free. This is shown by the fact that it is not the will he wants. It is also true, though in a different way, that the will of the wanton addict is not free. The wanton addict neither has the will he wants nor has a will that differs from the will he wants. Since he has no volitions of the second order, the freedom of his will cannot be a problem for him. He lacks it, so to speak, by default.
>
> (Frankfurt 1971, 22)

The unwilling addict presents conflicts at a higher level with what she wishes to be: her action of taking the drug does not issue from a second-order volition which is aligned with her will. The reason why the unwilling addict is not able to exercise free will, however, is not the obvious circumstance that her will is not free, that she is not free to want what she wants to want (she is an addict), but merely that there is a *discrepancy* between her will and her second-order volitions, that is, she does not have the will she wants.

In this passage, then, Frankfurt seems to favour concept (b) of free will (cf. section 1.4), that is, that a person exercises free will if her will and her second-order volitions conform. In order for the agent to assure that the will she has is the will she wants to have, a simple relation of *accord* between higher and lower-order wants is required.

Accordingly, it is possible to provide a third definition of freedom of the will:

> *(Def. 3) Free will as mere conformity*. A person exercises free will when she has the will she wants, even if she has no causal control over her will.

What is, then, the grounding element for freedom according to *(Def. 3)?* It seems that with *(Def. 3)* Frankfurt is requiring neither any form of counterfactual control between higher-level volitions and lower-level desires, nor any particular causal relation between the two levels, i.e. he is not requiring that the higher-level motivational elements causally produce the lower-level ones. For this reason, neither of the two elements grounding *(Def. 1)* and *(Def. 2)* seem to have a place here. It seems, then, that freedom as a mere conformity, according to which free will only requires the accord between volition of

different orders, lacks as such a proper grounding element. Indeed, nothing in *(Def.3)* explains or motivate why a simple relation of accord between the different levels of a person's will is able to secure freedom of the will. In short, as it stands, *(Def. 3)* is not sufficient to ground free will.[20]

Let us now sum up the three different conceptions of free will defined above. To exercise freedom of the will according to *(Def. 3)* requires the acceptance of only weak criteria. In order to exercise free will the agent has to:

I. have second-order volitions,
II. not have first-order volitions that are not in accord with those second-order volitions.

On the other hand, to exercise free will according to *(Def. 1)* and *(Def. 2)* seems to require the acceptance of further, stronger criteria (cf. Stump 1988). In addition to (I) and (II), the agent has to:

III. have the first-order volitions she has *because of* her second-order volitions, that is, her second-order volitions have produced his first-order volitions *(Def. 1* and *Def. 2).*
IV. and if her second-order volitions had been different, she would have had different first-order volitions *(Def. 1).*

In the next sections, I will suggest some reasons why such different conceptions of free will emerged in Frankfurt's theory (particularly, section 1.6). Moreover, I will make use of this distinction to offer an original formulation of the problem of the willing addict (section 1.5).[21] In the very last section, I will also apply this distinction to revisit Watson's (1975) classical objection to Frankfurt's theory (section 1.7).

[20] Later on, Frankfurt provides this conception of free will with a possible grounding element: the one of *wholehearted satisfaction* (cf. Frankfurt 1992). In chapter 2, section 2.5, I argue that such a ground is not effective enough to motivate an accord theory of free will.

[21] In chapter 3, section 3.8, I offer as well an original solution to the problem of the willing's addict freedom. There, I suggest that the willing addict is only allowed to exercise free will in a weaker sense than the one claimed by Frankfurt (1971).

1.5 The freedom of the willing addict

The tension between the three definitions of free will becomes more evident towards the end of the paper (§IV), when Frankfurt introduces the analysis of the willing addict. The willing addict has conflicting first-order desires with regard to taking a drug. But the second-order volitions of the willing addict are in line with her addictive first-order desire to take the drug: she wants to take the drug and wants her will to be formed by such a desire. According to Frankfurt, then:

> The willing addict's *will is not free*, for his desire to take the drug will be effective regardless of whether or not he wants this desire to constitute his will. But when he takes the drug, he *takes it freely* and of *his own free will*.
>
> (Frankfurt 1971, 24, *my emphasis*)

At a first glance, the understanding of such a case is not completely clear. Indeed, even if the willing addict's will is not free, she *does* have the will she wants to have. In light of this, how is it possible for a person to do something of her "own free will" if her "will is not free"? The case of the freedom of the willing addict is very controversial.[22] However, I suggest analysing her case in terms of the three definitions of free will provided above.

According to *(Def. 3)*, that is *free will as mere conformity*, the willing addict is accorded freedom of the will. For the will of the willing addict *conforms* to her second-order volitions: she wants to take the drug and wants her will to be formed by such a desire.

According to *(Def. 1)*, that is *free will as counterfactual causation*, the willing addict fails to exercise free will, despite the fact that her will is precisely what she wants it to be. More precisely, to exercise freedom of the will according to *(Def. 1)*, the willing addict should be able to satisfy both *actual* and *counterfactual control* over her will. As a matter of fact, the willing addict fails to satisfy the condition of counterfactual control, which is essential for satisfying *(Def. 1)*. The willing addict does not possess counterfactual control over her will; that is, she could not have done otherwise than constitute the will as she did, because her

[22] See for useful remarks on this issue Neely (1974); Locke (1975); Slote (1980); Taylor (2005).

will is subjected to an inescapable physiological addiction. This is sufficient to say that the willing addict fails to satisfy *(Def. 1)*.

According to *(Def. 2)*, however, the willing addict should only satisfy *actual causation*. That is, her higher-level motivational elements should be in a direct, internal relation with her lower-level ones. At first glance, it seems that here the direct relation between higher-level desires and lower-level desires is not respected, because the willing addict's desire is effective due to her physiological addiction. In the light of this, if the willing addict's first-order desire for the drug is due to an addiction, it is not correct to say that her higher-level desires causally produced her lower-level ones.

However, in the passage quoted above, Frankfurt seems to claim the opposite, that is, that when the willing addict takes the drug, she "*takes it freely* and of *her own free will.*" Now, Frankfurt is clearly pointing out that the willing addict, compared to the unwilling addict, exercises freedom of the will. Indeed, to better illustrate the case of the willing addict, Frankfurt is inclined to extend the kind of freedom she enjoys to other, non-addicted persons:

> Suppose that a person has done what he wanted to do, that he did it *because he wanted to do it*, and that the will by which he was moved when he did it was his will because it was the will he wanted. Then *he did it freely* and *of his own free will*. Even supposing that he could have done otherwise, he would not have done otherwise; and even supposing that he could have had a different will, he would not have wanted his will to differ from what it was.
>
> (Frankfurt 1971, 24, *my emphasis*)

The agent in Frankfurt's example exercises free will because she respects *(Def. 2)*, that is a condition of actual control on her will. Accordingly, the only way to say that the willing addict exercises free will in the same way is to claim that the willing addict's second-level wants *do* causally produce the first-level ones, i.e. she is free in the actual scenario to want what she wants to want. However, to accept this conclusion one has to postulate, as Frankfurt does, an *overdetermination* of the willing addict's first-order desires to take the drug: her desire is effective because she is addicted to the drug but, at the same time, it is her effective desire because she wants it to be (Frankfurt 1971, 25).

If, however, one wants to reject the overdetermination strategy, one has to conclude in contrast with Frankfurt's position that the willing addict fails to exercise free will even according to *(Def. 2)*, that is, *free will as actual causation*. Accordingly, the only way in which the willing addict is allowed to exercise free will would be by *(Def. 3)*, that is, free will as a mere conformity between higher and lower-level desires.

1.6 Freedom as a twofold ability

I suggest that the tension between the three concepts of free will arises because of Frankfurt's different treatment of the problem of free will and of the one of moral responsibility (1971, 14-16). According to Frankfurt:

> It is a vexed question just how "he could have done otherwise" is to be understood in contexts such as this one. But although this question is important to the theory of freedom, it has no bearing on the theory of moral responsibility. For the assumption that a person is morally responsible for what he has done does not entail that the person was in a position to have whatever will he wanted. This assumption *does* entail that the person did what he did freely, or that he did it of his own free will. It is a mistake, however, to believe that someone acts freely only when he is free to do whatever he wants or that he acts of his own free will only if his will is free.
>
> (Frankfurt 1971, 24)

In this passage, Frankfurt attempts to spell out the pertinence of the ability to do otherwise in contexts that concerns both freedom of the will and moral responsibility. Frankfurt thinks that questions about the ability to do otherwise are irrelevant when accounting for moral responsibility (Frankfurt 1969).[23] This means that a person can be morally responsible for what she has done even if

[23] Frankfurt (1969) imagines cases in which an agent acts freely but at the same time she is not able to do otherwise. He imagines a situation in which an external intervention - say, an evil neuroscientist - manipulates the agent's brain, and blocks an alternative course of action only in those situations where the agent decides to act contrary to the neuroscientist's desires. However, the agent happens to act in accordance with the neuroscientist's desires autonomously, without any need for the neuroscientist to intervene. In such a case, the actual course of action is a free one, and the agent is responsible for her choice in performing that action, even if he was not really able to do otherwise at the moment of that action.

she was not able to do otherwise, that is, even if she cannot *have had whatever will she wanted*. However, according to Frankfurt, to be morally responsible entails that a person did what she did *of her own free will*. As an explanation of the distinction between 'having the will one wants' and 'acting on one's own free will,' Frankfurt claims that it is a mistake to believe that someone acts of her own free will *only* if her will is free. For Frankfurt, a person is also free if she acts on her own free will. That is, a person is free if she acts with a will that is free *and* if she acts of her own free will.

To clarify these quite obscure distinctions, I propose to rephrase Frankfurt's claims about the difference between free will and moral responsibility in terms of the three definitions of free will provided above.

According to Frankfurt, to be morally responsible it is sufficient that a person enjoys free will in the sense described by *(Def.2),* that is, *free will as actual causation*: if the person's will aligns with her second-order volitions, and her second-order volitions are internally causally related to her lower-level ones, she is acting of her own free will and she is morally responsible for what she has done, even if she could not have done otherwise and have another will instead. In Frankfurt's words:

> Moreover, since the will that moved him when he acted was his will *because he wanted it to be*, he cannot claim that his will was forced upon him or that he was a passive bystander to its constitution. Under these conditions, it is quite irrelevant to the evaluation of his moral responsibility to inquire whether the alternatives that he opted against were actually available to him.
>
> (Frankfurt 1971, 24, *my emphasis*)

The exercise of free will, however, is a more complex matter. Frankfurt says that the conditional ability to do otherwise is important for accounting for free will. That is, he seems to favor *(Def. 1)* of free will: free will as *counterfactual causation*, according to which a person exercises free will if she is free to have the will she wants and she could have constituted her will otherwise than she did. However, he claims as well that only *(Def. 2),* that is, *actual causation*, is necessary to exercise free will: in the passage quoted above, the agent is morally responsible for her actions exactly because she is acting of her own free will.

The best way to understand this distinction is to say that Frankfurt seems to point to *two* different abilities involved in the exercise of free will:

i. Acting of one' s own free will
ii. Acting with a will that is free

At the same time, he seems to say that while (i), 'acting with one's own free will', only requires internal control *(Def. 2),* (ii), 'acting with a will that is free', requires counterfactual control as well *(Def. 1).*

However, what remains outside of this analysis of free will is the case of the *willing addict*. As I argued in the previous section, if one rejects the overdetermination strategy, then the willing addict is exercising freedom of the will neither in the sense of 'acting with a will that is free', nor in the weaker sense of 'acting of one's own free will'. If my analysis is correct, the willing addict is allowed to enjoy free will only according to *(Def. 3),* that is, free will as a mere accord between higher and lower-order volitions. Accordingly, the only way Frankfurt has to save the analysis of the willing addict as an agent who "act freely" and "of his own free will" would be the one of dropping even the assumption of *actual causation*, and to claim that to exercise (i), that is, to 'act with one's own free will', the agent has only to satisfy *(Def.3),* that is, one has only to have higher-order volitions in accord with the lower-level ones.

In the rest of the dissertation, I will refer only to meaning (i) of free will, that is, I understand free will as the ability to 'act with one's own free will.' My aim is to show that a coherent account of free will so understood necessarily drops the condition of internal causal control. Particularly, in chapter 2, I will argue that after his seminal paper Frankfurt tries to reinforce an account of free will as *(Def.2)* with different concepts (i.e., *decision*). However, I show that maintaining this condition, as Frankfurt tries to do, would only lead to contradictory intuitions about the ability to exercise free will.

1.7 Watson's criticism revisited

In the previous sections, I argued that Frankfurt's (1971) hierarchical theory is ambiguous between three different conceptions of free will. I also suggested that such an ambiguity lies in the different roles played by higher-order volitions in Frankfurt's theory of free will. At the same time, I argued that the three definitions of free will discussed above provide Frankfurt's theory with different possible grounding elements for freedom. However, it seems that there is, *prima facie*, a preliminary structural problem with Frankfurt's theory. A well-known objection, firstly presented by Watson (1975), casts doubts on the very role of higher-order volitions as a securing element for freedom of the will. In this section I aim to show that the three different definitions of free will provided above account for different strategies of reply to Watson's concern. Nonetheless, I conclude that none of the strategies is able to justify the role of higher-order volitions as a grounding element for freedom of the will.

Watson's concern is that the *hierarchical structure* of the will favoured by Frankfurt, and particularly the role of higher-order volitions, is not adequate to secure free will and self-determination. According to Watson, Frankfurt's use of a hierarchical structure to guarantee free will is meant to account for the agent's self-determination with respect to her own desires. The hierarchical structure supports the intuitive claim that some desires are more truly 'the agent's own' than other desires. Even if all desires are in a sense the agent's own, nonetheless the agent usually identifies herself with some particular wants - in Frankfurt's terms, 'higher-order volitions' - and such wants are considered by the agent herself as the expression of her 'real self.' In this light, in the case of conflicts in the agent's motivational system, Frankfurt (1971) claims that the first-order (or second-order) desire which is in the end endorsed by the agent is the one which presents a higher-order volition concerning it, while the other desire in conflict only has a volition of a lower-level order concerning it.

However, Watson contends, this condition is enough to guarantee free will only if the higher-order volition of the agent is *itself* one by which the agent really wants to be determined. In other words, if it is itself freely willed. As Watson suggests, though, Frankfurt's theory does not seem to have the

resources to justify this condition: according to the theory as it stands, there is apparently no reason to give a special status in the hierarchy to higher-order volitions rather than to the lower-level ones. Since higher-order volitions are just desires, there is no immediate explanation for the claim that it is *their* role in the agent's motivational structure that assures that the agent's desires are freely willed.

More precisely, Watson points out to two problems. On the one hand, to say that freedom of a higher-order volition is guaranteed by forming another volition of an even higher-order would lead to a *regress*, and to terminate it at any particular level would be quite arbitrary. Let us call this issue the *Regress Problem*.

On the other hand, besides the *Regress Problem*, there is apparently no reason in Frankfurt's theory according to which an agent should "care about one's higher-order volitions" more than about lower-order volitions. Let us call this issue the *Specialness Problem*. In Watson's words:

> Can't one be a wanton, so to speak, with respect to one's second-order desires and volitions?
>
> (Watson 1975, 217)

And moreover:

> Higher-order volitions are just, after all, desires, and nothing about their level gives them any special authority with respect to externality. If they have authority, they are *given* it by something else. To have significance, the hierarchy must be grounded in something else that precludes externality.
>
> (Watson 1987, 149)

It is Frankfurt's idea that the agent exercises free will if and only if the desires which form the agent's will are *internal* and not alien to the self. If, however, the agent's volitions are *external* to the self, then the agent is not

exercising free will (cf. Frankfurt 1977).[24] Nevertheless, if this is true, what can prevent higher-order desires from being themselves external to the self? According to Watson, there is apparently no reason in Frankfurt's theory why higher-order volitions can prevent the risk of externality in a different way from lower-order volitions. Crucially, Watson claims, it is impossible to prevent the risk of externality for second-order volitions as well. Indeed, since second-order volitions are simply desires, it is impossible to prevent them from being external to the self by just pointing out to their position in the hierarchy.

Watson himself, for example, thinks that externality is prevented only if the hierarchy of desires is grounded in our *evaluations*. According to Watson, an addict who is lead to the action of taking a drug by a desire that is rejected by her "valuational system" is not, for this reason, identifying with such a desire (28-29). However, Watson gives up on this account in 1987. The reason for this, Watson says, is that it is always possible for an agent to identify with a course of action which she does not consider the best, or the one that matters most.[25]

How Frankfurt's (1971) theory is affected by this objection? In the previous sections, I argued that in Frankfurt (1971) higher-order volitions perform different roles when accounting for free will. More precisely, I argued that the *condition of control* which is supposed to link the agent's higher-order motivational elements with the lower-order ones is spelled out by Frankfurt in different ways, thus giving rise to three definitions of free will. In light of this, I propose to rephrase Watson's objection *(WO)* by focusing on the different role which higher-order volitions perform in the different definitions.

On the one hand, if one considers *(Def. 1)* and *(Def.2), (WO)* to the notion of higher-order volitions can be expressed as follows:

[24] In the next chapter, section 2.2, I will take into account in details the issue of internality/externality.

[25] Another reading of the role of second-order volitions in the agent's motivational structure is advanced by Eleonore Stump. She claims that second-order desires secure freedom not just because they are of a higher-order, but "because the agent's second-order desires are the expressions of his intellect's reflection on his will" (Stump 1988, 408).

(WO/causal): how does the fact that higher-order volitions causally produce lower-level desires reply to the *Regress Problem* and to the *Specialness Problem*?

On the other hand, if one considers *(Def.3)*, *(WO)* can be expressed as follows:

(WO/accord): how does the fact that higher-order volitions are in accord with lower-order reply to the *Regress Problem* and to the *Specialness Problem*?

I begin by considering *(WO/causal)*. One way to explain why, in this picture, second-order volitions are themselves free, and thus are able to ground freedom for lower-level desires, is to say that the second-order volition is free because it can always be regarded as free from the standpoint of an even higher-order volition. As Watson notes, however, such a reply runs the risk of incur in the *Regress Problem*.

By making use of *(Def.1)* or *(Def.2)*, such a claim can be rephrased by stating that the freedom of every higher-order volition is guaranteed by the fact that it is causally produced by a further free higher-order volition, and so on ad infinitum. In other words, *(Def. 1)* and *(Def.2)* work by extending causal control all the way up to the hierarchy, as it is expressed by the following condition:

(C) To will freely, the agent has to have causal control on her volitions all the way up to the hierarchy.

Condition (C) is, however, problematic in at least two respects. On the one hand, it is not clear whether infinite hierarchical structures are able to prevent the risk of regress pointed out by Watson. On the other hand, it seems that, even in their causal form, infinite motivational structures are not able to avoid the risk of externality for the agent's higher-order volitions.

According to Lehrer (1980), for example, infinite causal hierarchical structures are adequate to reply to the *Regress Problem*. On his account, the regress exists but is not vicious, because when a person acts freely, she possibly possesses controlling preferences of this sort all the way up (cf. Lehrer 1980,

193). According to Lehrer, infinite hierarchical structures are plausible if one claims that *indifference*, too, is an attitude: "if *S* freely prefers *K* at level *n*, then S must either prefer to have that preference or must be indifferent" (193). In this light, to be indifferent is a "primitive" preference and there are no issues in the claim that people have preferences all the way up. Indeed, starting from a certain level, people are indifferent all the rest of the way up. Lehrer suggests as well that even from a psychological point of view, infinite hierarchical structures are not a problem, because it is possible for persons to process things computationally, a step at a time, without grasping the whole sequence in a single moment of intuition.

Such a strategy, however, seems not satisfactory. On the one hand, as Shatz (1985) suggests, the condition of control offered by indifference is "entirely negative; that is, 'control' is being exercised by the absence of a specific preference" (457). On the other hand, such a reply seems to be incompatibilist in its essence. Indeed, if every level of the hierarchy is subjected to the agent's internal and causal control, then this excludes the possibility that at any level the agent's wants are caused by an *external cause*. Zimmerman (1981, 359), for example, is well-aware of such an issue when he writes:

> It is no part of this program to claim that we are always free 'to please as we please as we please'… Precisely because he concedes the truth of determinism the compatibilist is perfectly willing to acknowledge that even where a desire is free, there is some point in the motivational hierarchy where the higher-order desire playing the crucial endorsing role is itself an unwilled, unendorsed part of the agent's motivational equipment, to be explained in terms of non-motivational causes.

Moreover, it seems that there is a further issue with condition (C), as it seems not sufficient to prevent the risk of externality. As Van Inwagen (1983) and Segerberg (1983) have pointed out, it might be that some external manipulator controls what an agent desires at all levels and, at the same time, what the agent would desire at all levels if her preferences were different. In such a scenario, all the agent's preferences are not in her control, but in the manipulator's control. In other words, in a manipulation-scenario, the mere

truth of (C) does not exclude the possibility of manipulation on the agent's motivational structure.[26] The risk of externality is then not avoided.

Does the use of *(Def. 1)* or *(Def. 2)* help reply to the *Specialness Problem?* To repeat, the problem, as Watson points out, is that there is apparently no reason in Frankfurt's theory why an agent should care about one's higher-order volitions more than about lower-order volitions. In this light, Watson contends, Frankfurt theory as it is does not elucidate "why or how a particular want may have, among all of a person's desires, the special property of being peculiarly his own" (Watson 1975, 218-219)

With *(Def.1)* or *(Def. 2),* one can claim that higher-order volitions are special because they causally produce lower-level volitions. However, this reply seems to just beg the question. Indeed, the agent may not care at all about the fact that higher-order volitions produce her will. At the same time, to claim that the freedom of every higher-order volition is guaranteed by the fact that it is causally produced by a further free higher-order volition does not say anything about *why or how* it is the higher-order volitions' role which grounds freedom. Without a plausible, further explanation for Frankfurt's claim, then, it seems that higher-order volitions are not sufficient to ground freedom for the hierarchy.

Let us now consider *(WO/accord)*. If one takes into account *(Def. 3)*, one has to explain why the *mere accord* between higher-order volitions and lower-order desires guarantees freedom for higher-level desires. Moreover, by doing so, one has to provide a reply to Watson's *Regress Problem* and *Specialness Problem*.

I begin by considering the *Regress Problem*. In the picture suggested by *(Def. 3),* to regard a desire as external is to have a contrary higher-order volition concerning it. In this light, second-order volitions are themselves free from externality if they do not have any contrary volition regarding them. However, even in this scenario, a higher-order volition can be so considered only from the standpoint of an even higher-order volition, and it is always the highest-order volition that is decisive. Does such a scenario lead to a regress?

It seems to me that a relation of mere accord between the levels of the hierarchy is not sufficient to stop the regress or to show that the regress, even

[26] For different perspectives on the manipulation-objection see Fischer (1994), Kane, (2002b), Pereboom (2001) and van Inwagen (1983).

if present, is not vicious. In this case, indeed, the presence of a regress prevents the determination of freedom for higher-level volitions, by preventing the agent's control with respect to her own desires. While with *(Def. 1)* or *(Def.2)* the entire motivational structure of the agent is put under her internal causal control, with *(Def. 3)* there is no assurance that the higher-order volitions of an agent are themselves internal to the agent's motivational structure: since they are simply desires, they do not apparently possess a particular status against externality.

Let us now consider the *Specialness Problem*. Frankfurt can point out that what in this picture assures that the volitions endorsed by the agent are themselves free is that they represent the 'true self' of the agent: only when the agent endorses volitions which are part of her real self, then she is acting of her own free will. However, this strategy seems to just beg the question, because there is no obvious reason for the claim that the higher-order volitions characterise the real self of the agent. In other words, it seems that without a proper 'theory of personality,' which could justify the claim that the higher-level self is the real self, such a claim is unwarranted (cf. for this objection Shatz 1985, Berofsky 1983).

In conclusion, nothing in *(Def. 3)* explains or motivate why a simple relation of accord between the different levels of a person's will is able to secure freedom of the will. If one accepts *(Def. 3)* of free will, then, it is not apparently possible to reply to Watson's objection by simply putting the burden of securing freedom on hierarchical structures and higher-order volitions.

1.8 Summary of the chapter

In this chapter, I took into account Frankfurt's classical theory of free will. I argued that Frankfurt's theory can be interpreted as providing an original grounding element for freedom, i.e. that free will is mostly grounded in the causal role of higher-order desires in the hierarchy. I suggested that Frankfurt defines a different ground for freedom by tightening the *conditions of control* of the agent on her desires so as to ensure that the agent controls her motivational structure through her higher-order volitions. However, I showed that this condition of control is not univocally spelled, because higher-order volitions

perform different roles in Frankfurt's account. As a consequence, I showed that Frankfurt's original theory is ambiguous between three different definitions of free will, which I called *(Def.1)*, *(Def.2)* and *(Def. 3)*. As a direct consequence of my claims, I argued that Frankfurt's theory provides different and unclear elements for grounding freedom.

The primary result of this chapter is then the suggestion that Frankfurt's original theory provides neither a single definition of free will, nor a univocal grounding element for freedom. Indeed, because of such an ambiguity, I argued that it is not clear if the role of higher-level desires is successful in the task of grounding freedom. The line of argument that I presented is, in this sense, quite different from classical Watson-style objections to Frankfurt's theory of free will.

Distinguishing between these three concepts of freedom served a number of additional purposes in this chapter. Specifically, such a distinction allowed me to offer an original formulation of the problem of the willing addict (section 1.5). Contrary to Frankfurt's claims, I showed that the willing addict does not exercise free will according to an actual causal conception of freedom, i.e. *(Def.2)*, but only following a mere conformity conception of it, i.e. *(Def. 3)*.

Moreover, I made use of this distinction to revisit Watson's (1975) classical objection to Frankfurt's theory (section 1.7). My conclusion is that Frankfurt's theory, even if enhanced with the distinction between three different conceptions of free will, is not able to justify the role of higher-order volitions as a grounding element for freedom of the will.

2

Freedom and Identification

2.1 Outline of the chapter

In this chapter, I discuss Frankfurt's claim that identification is sufficient to secure freedom. I argue that Frankfurt makes use of two different notions of identification to support his claim. In section 2.3 and 2.4, I show that Frankfurt argues for identification by grounding it in the notion of *decision*. I call such a notion *active identification (AI)* and I contend that it does not work to ground freedom. In sections 2.5 and 2.6, I show that Frankfurt rejects this account and argues instead for a notion of wholehearted identification secured by the concept of *satisfaction*. I call such a notion *wholehearted identification (WI)* and I suggest that this approach is also doomed to failure. My conclusion, then, is that Frankfurt's notion of identification is not sufficient to ground his hierarchical theory of freedom.

2.2 Identification and Externality

In chapter 1, I revisited Watson's objection about the role of higher-order volitions as a securing element for freedom of the will. There, I showed that the three different definitions of free will present in Frankfurt (1971) provide different strategies of reply to Watson's concerns about the *Regress Problem* and the *Speciality Problem*. Nonetheless, I concluded that none of the strategies is able to justify the role of higher-order volitions as a grounding element for freedom of the will. In this chapter, I focus on another aspect of Watson's objection, that I call the *Externality argument*. Particularly, I investigate Frankfurt's notion of *identification*, developed in various directions after his 1971 paper, and I show that such a notion is exactly an attempt to reply to this argument.

Indeed, one way to understand Watson's objection is by focusing on the notion of *externality* (Watson 1987). What the *Externality argument* claims is that an agent is free if and only if she is able to act on *her* wants and desires. If the desires on which she acts are, on the contrary, alien or external to her motivational structure, then she is not acting with a free will.

Such an argument is mostly supported by compatibilist accounts of free will, especially classical compatibilism (cf. Shatz 1985). In this regard, classical compatibilism considers free agency in terms of 'external' impediments to the will of the agent, where external is meant to indicate impediments that are found outside the body of the agent. However, the main problem with such an account lies exactly in its 'thin' notion of will, and in the consequent lack of focus on more 'internal' impediments, such as addiction, compulsion, phobia and the like. What strikes us as problematic in the classic compatibilist account is in fact the absence of a satisfactory criterion for internality. Compatibilist theories, indeed, seem unable to mark a distinction between externality and internality with regard to one's desires and volitions. The consequence of this failure is an account of freedom that is too thin to provide a realistic criterion for freedom of the will.

Frankfurt's hierarchical theory seems to deal with these problems in a better way. The general strategy he adopts is one of construing the agent's

motivational organisation as composed of different levels of desires. In this sense, Frankfurt shows that deviant desires or intentions (e.g. the desires of an addicted agent) are the result of an incorrect mesh between different levels of the person's motivational structure. Frankfurt also links freedom to the notion of what is internal to the agent's motivational structure. However, the criterion that is supposed to mark internality is not merely physical, but is instead related to the absence of feelings of estrangement from the agent's point of view.

As noted by Velleman (2002), Frankfurt's account of freedom might be described as an attempt to find criteria for establishing which psychological elements are internal, and which are external to the self. The reason for this, Frankfurt says, is that an agent may sometimes consider a desire as 'external' to her motivation, and so alien to herself, even if such a desire is in a basic sense 'her desire'. In turn, such feeling of estrangement is usually connected to the agent's rejection of her desire as a free and autonomous happening. The desires associated with such a feeling of estrangement are usually considered as unfree, and the agent is just a "passive bystander" to their presence in her motivation (cf. Frankfurt 1977).

To see why the concept of externality is important, consider the possibility of conflicts between the different levels of the agent's will. The hierarchical structure of the will proposed by Frankfurt can account for at least two kinds of conflict within it. On the one hand, "there may be a conflict between how someone wants to be motivated and the desire by which he is in fact most powerfully moved" (Frankfurt 1987, 164). In this case, the conflict emerges between what the agent really wants (e.g. to refrain from smoking), and the desire that turns out in the end to be the strongest one, that is, the desire which effectively forms the will of the agent (e.g. to smoke). Since the agent is not able to put in alignment her first-order desires with her higher-order volitions, the resultant action is not free, because it stems from a desire that is not controlled by the agent herself.

On the other hand, there can be "a lack of coherence within the realm of the person's higher-order volitions themselves" (165). Because they are desires, higher-order volitions can conflict with each other. For example, an agent may be divided between a "jealously spiteful desire to injure" a friend and a desire to suppress this motive and benefit her friend (Frankfurt 1977, 63). In this case,

the agent is divided about what she really wants, that is, she is uncertain about which one of her second-order volitions should produce her effective will. In such a case, it is possible to imagine that a person may have volitions of a still higher order to solve the conflict between second-order volitions. ("There is no theoretical limit to the length of the series of desires of higher and higher orders", Frankfurt 1971, 21).

In both cases, the conflicting desires are perceived by the agent as external to her motivational structure. What, in such cases, can prevent the agent from obstinately refusing to endorse one of her higher-order volitions, so as to act with a will that she wants to have, a will which is not 'external' to her volitional self?

In the attempt of addressing these issues, Frankfurt (1971) makes use of the hierarchical structure of the will to account for the agent's self-determination with respect to her own desires: when the agent's desires are endorsed by a higher-order volition, they can be considered as internal to her motivation, thus securing freedom of the will. Moreover, as argued in chapter 1, Frankfurt's (1971) theory of free will finds a viable ground for freedom not just by endorsing higher-order volitions, but by establishing a condition of control from higher-order volitions to lower-level ones. Such a condition can be said to distinguish between desires that are internal to the agent, because they are brought about by higher-order volitions, and desires that are external, because they are not brought about by higher-order volitions.

However, toward the end of the paper, Frankfurt introduces a new concept to further justify his claims: *identification*. In Frankfurt's words:

> When a person *identifies himself decisively* with one of his first-order desires, this commitment "resounds" throughout the potentially endless array of higher orders. Consider a person who, without reservation or conflict, wants to be motivated by the desire to concentrate on his work. The fact that his second-order volition to be moved by this desire is a decisive one means that there is no room for questions concerning the pertinence of volition of higher orders.
>
> (Frankfurt 1971, 21, *my emphasis*)

Roughly, identification is a psychic process by which the agent's motivational elements, in particular her desires, become in some sense 'her

own'. According to such a view, when an agent identifies with a desire, it becomes internal to her will, and when she rejects that desire, it becomes external. For Frankfurt, when the agent acts on a desire with which she is identified, she is acting as an active and autonomous agent. On the contrary, when she acts on desires with which she does not identify decisively, she is acting in a passive way, because she is moved by desires that she considers as alien to herself. To exercise free will is then not just reducible to the agent's correct mesh between volitions and will, but instead it is now a matter of *decisive identification with certain desires*, while a lack of such identification would entail the absence of free will. When the agent acts on desires with which she has identified, her will becomes constituted by the volitions aligned with her identification, that is, with the will she wants to have at a higher-order of reflection.[27]

The difference between the two solutions proposed by Frankfurt to the problem of conflicts in the agent's motivation is well exemplified by Bratman (2002). Bratman notes that in Frankfurt's (1971) there is a tension between two different views. One view is that identification can be reduced to the state of affairs in which one's relevant highest-order volition favours one's being moved by that desire. The second view suggests that identification with a first-order desire involves a highest-order volition in its favour but is not simply reducible to such a highest-order volition.

The former view, however, proved to be wrong. As discussed in chapter 1, there is a reasonable doubt that the role of higher-order volitions as such is adequate to ground free will.[28] Particularly, Watson's (1975) objection forces Frankfurt to reject the view that externality is simply avoided by the endorsing or even controlling higher-order volitions of the agent.[29]

Let us now consider the latter view. In such a picture, it is no longer the concept of a higher-order volition that is doing the job of defining internality and grounding freedom. Rather, such a work is now done by a further element,

[27] A different definition of the notion of identification is given, for example, by Alfred Mele: "to identify with a desire one has is, roughly, a matter of desiring to continue to have that desire and believing that one's having it is a good thing" (Mele 2003, 227 and Mele 1995, 117).

[28] Susan Wolf points out a similar concern, when she writes that a person who is alienated from a first-order choice "can be alienated from her higher-order choices as well" (Wolf 1990, 30).

[29] As Bratman notes, a second-order desire is just "one more desire in the psychic stew" (Bratman 2007, 23).

that is, the agent's acts of identification or decisive commitment. When an agent "*identifies … decisively* with one of his first-order desires", Frankfurt claims, she is securing freedom for the higher-order volition concerning that desire. Unfortunately, however, Frankfurt (1971) does not provide any particular argument to justify this claim.

Such a lack of arguments brings about immediate problems. Indeed, it seems that the notion of identification as a higher-order phenomenon which marks a criterion for internality and, consequently, secures freedom needs a proper basis to accomplish the task. Watson (1975), (1987) and Bratman (2002), among others, have raised such a concern about Frankfurt's use of the notion of identification. Roughly, the worry is that identification as Frankfurt characterizes it in 1971 is not at all sufficient to protect the agent's will from the risk of externality. In parallel with the criticism to the notion of higher-order desire, it seems that there is nothing in the very concept of decisive identification which is able to explain why, when an agent identifies with some of her desires, she is securing freedom for higher-order volitions. In other words, the concern is that the notion of identification as elucidated by Frankfurt does not sufficiently underpin a hierarchical account of freedom.

In the remainder of this chapter, I show that Frankfurt attempts to motivate identification in his later papers (particularly, 1987 and 1992) by taking into account the second view suggested by Bratman (2002)[30] - namely that identification with a first-order desire involves a highest-order volition in its favour but is not simply reducible to such a highest-order volition - and by adopting various new concepts as proper grounds to justify his position. Particularly, I argue that Frankfurt makes use of two different notions of identification in order to defend his account of freedom. In section 2.3 and 2.4, I show that Frankfurt argues for identification by grounding it in the notion of decision. I call such a notion *active identification (AI)* and I contend that it does not work to ground freedom. In sections 2.5 and 2.6, I show that Frankfurt rejects this account and argues instead for a notion of wholehearted identification secured by the concept of satisfaction. I call such a notion *wholehearted identification (WI)* and I suggest that this approach is also doomed

[30] Cf. above, 42.

to failure. My conclusion, then, is that Frankfurt's notion of identification is not sufficient to ground his hierarchical theory of freedom.

2.3 Identification as Decision

In this section, I examine the first amendment to the notion of identification presented in Frankfurt (1987). In this paper, Frankfurt proposes to validate second-order volitions by strengthening the role of decisive identification, which in turn is justified by the notion of *deliberate decision* with respect to one's own desires. My aim is to show that such an attempt fails, and that the concerns about the role of identification as used by Frankfurt in his original theory of freedom are not addressed in an effective way in this later paper.

Frankfurt (1987) aims to justify the claim that "making a decisive commitment" to avoid an interminable regress of higher-order volitions is not, as Watson's objection argues, unacceptably arbitrary. The overall strategy presented by Frankfurt stems from a comparison between the agent's evaluation of her motivational structure and the practice of arithmetic. In this context, Frankfurt makes use of the notion of *deliberate decision* with respect to one's own desires to show that identification helps to avoid externality because it is secured by the agent's deliberate decision about her motivational structure.

Frankfurt suggests that the situation of ambivalence with respect to one's own desires is significantly similar to the process of arithmetical calculation. When calculation begins different answers to a problem are possible. However, at the point of terminating a calculation the possible answers are reduced to one, and one consequently has no reasons to carry on the calculation. As with the case of arithmetic, a person's decisive identification with a second-order volition essentially consists in the person's persuasion that a further examination of her motivational situation will nonetheless lead to the same result. In other words, even in the case of a further scrutiny, any higher-order volition of the agent will point out to the agent's preference for a specific first-order desire, the one recommended by the second-order volition. Therefore, for the agent to engage in a potentially endless sequence of evaluations is pointless, exactly because the person has decisively identified herself with that particular second-order volition.

While certainly evocative, Frankfurt calculation metaphor seems to provide a quite obscure explanation about the role of decisive identification in Frankfurt's picture of freedom. My argument is that such an obscurity is a consequence of the presence of an ambiguity in the notion of decisive identification in Frankfurt (1987). To illustrate my claim, let me quote Frankfurt at length here:

> Both in the case of desires and in the case of arithmetic a person can without arbitrariness terminate a potentially endless sequence of evaluations *when he finds that there is no disturbing conflict*, either between results already obtained or between a result already obtained and one he might reasonably expect to obtain if the sequence were to continue. Terminating the sequence at that point - the point at which there is no conflict or doubt - is not arbitrary. For the only reason to continue the sequence would be to cope with an actual conflict or with the possibility that a conflict might occur. Given that the person does not have this reason to continue, it is hardly arbitrary for him to stop.
>
> (Frankfurt 1987, 169, my emphasis)

And moreover:

> On the other hand, a sequence of calculations might end because the person conducting it *decides for some reason* to adopt a certain result. It may be that he is unequivocally confident that this result is correct, and therefore believes that there is no use for further inquiry. Or perhaps he believes that even though there is some likelihood that the result is not correct, the cost to him of further inquiry - in time or in effort or in lost opportunities - is greater than the value to him of reducing the likelihood of error. In either event there may be a "decisive" identification on his part. In a sense that I shall endeavor to explain, such an identification resounds through an unlimited sequence of possible further reconsiderations of his decision.
>
> (Frankfurt 1987, 170, my emphasis)

I suggest that according to Frankfurt's metaphor the process of decisive identification comprises two different elements. Particularly, it seems that a person properly identifies with a lower-order desire in at least two ways:

(1) when she *has an unopposed higher-order volition* to act in accord with the lower-order desire, and she judges that any further deliberation would be pointless, and/or

(2) when she *decides* that a certain second-order volition is what she really wants to be her will.

To illustrate claim (1), let us consider again Frankfurt's arithmetical metaphor. As suggested by Frankfurt, when calculation begins different answers to a problem are possible. However, at a certain point the agent terminates a calculation because she "finds" that there are no more conflicts in the possible answers, and she consequently has no reasons to carry on with the calculation. When this happens, there is a decisive identification of the subject with the object of her calculation. Such identification "resounds through an unlimited sequence of possible further reconsiderations" of her decision (169). It resounds endlessly, and this is exactly the reason why the person stops the calculation.

Similarly, a person is moved to reflect on her own desires, either because there effectively is a conflict among them or because she has a lack of confidence in the fact that she is satisfied with her motives. However, as in the case of arithmetic, to terminate such a sequence of evaluation is not arbitrary. Indeed, according to Frankfurt, in such a situation too, a person who is engaged in a reflection about her own desires "finds" at a certain point that there are no more conflicts between her first-order desires and second-order volitions, or between her second-order volitions themselves. When she judges that there is no reason to carry on the scrutiny of her motives, she decisively identifies with the lower order desire which corresponds to an unopposed higher-order volition.

It seems appropriate to understand the process described by Frankfurt's claim (1) as a sort of *discovery* and *recognition* made by the person. When the agent recognises an unopposed higher-order volition in accord with a lower-order desire, she judges that any further deliberation would be unnecessary,

and she identifies herself with the lower-order desire in alignment with that unopposed high-order volition.[31]

In this picture, the person has no need to create a different alignment through the act of identification: some higher-order volitions are already in alignment with the agent's will, and when she discovers that this is the case, the agent secures her freedom and activity through her identification with them. However, it is exactly her act of *decisive identification and commitment* with her unopposed higher-order volitions that prevents externality for her motivational structure. The agent's decisive identification with some of her desires is what guarantees that the unopposed desires which occur in her, and so that are in a sense 'external' to her, become incorporated into her will, thus becoming 'internal' to her own will.

At this stage of the paper, Frankfurt is then proposing the following criterion for internality: higher-order volitions are internal to the motivational structure of the agent when they are *unopposed* or *decisive* with respect to the agent's will. Such a status, however, is fully ratified by the agent's endorsement by means of decisive identification. If this is Frankfurt's answer, however, then *prima facie* it is not very different from the original formulation of the notion of identification presented in Frankfurt (1971). In this picture, the burden of securing freedom is not just on the notion of unopposed higher-order volitions, but also on the agent's ability to identify with them. This claim, however, seems unjustified, and the risk of externality for unopposed volitions arises again. Indeed, the notion of decisive identification with unopposed higher-order volitions, in itself, does not suggest a plausible explanation for Frankfurt's claims.

At this point, Frankfurt can reply that in this picture identification is justified with the same arguments expressed in Frankfurt (1975): the freedom of the higher-order volitions, ratified by the act of decisive identification, is in the end simply *self-determining*; in this sense, higher-order volitions self-constitute the activity of the person and it is impossible for the agent to be a passive bystander to them. For this reason, questions about the legitimacy of identification simply cannot arise (Frankfurt 1975, 54). In this picture, the burden of the argument is once again on the notion of higher-order volitions.

[31] In section 2.4, I argue that a such a conception of *freedom as discovery* is what justifies Frankfurt's proposal of *wholehearted satisfaction*.

Unopposed higher-order volitions are free and exempt from externality simply because they are higher on the hierarchy than first-order desires, and because (in an admittedly mysterious sense) they self-constitute the activity of the agent. But if this is Frankfurt's characterisation of identification, then it runs the risk to fall again under Watson's primitive objection: higher-order volitions, since they are desires, are subjected to externality, and to claim that some of them are unopposed or decisive with respect to the agent's will by just pointing out their position in the hierarchy seems not to be a satisfactory reply.

I suspect that it is exactly due to the lack of a convincing argument for endorsing such a picture that Frankfurt is ultimately led to propose a further element to justify identification. Indeed, to avoid this kind of issues, Frankfurt introduces at the end of the paper the further notion of *decision*. A person properly identifies with a lower-order desire, Frankfurt says, when:

(2) she *decides* that a certain higher-order volition is what she really wants to be her will.

According to claim (2), what prevents an infinite regress of higher-order volitions and secures freedom for higher-order desires is located in the agent's *act of decision* (Frankfurt 1987, 170). Frankfurt says that the role of deliberate decision is one of "creating an intention" (172). Intentions, in this sense, involve a commitment to actively suppress some desires in favour of others. Accordingly, they suggest the possibility of changing the configuration of one's will: in deciding in favour of a desire, and in identifying with it, the agent is creating an intention that will enable her to reconfigure her will and to make the desire on which she decides fully her own. How does Frankfurt justify this claim?

Here, Frankfurt is suggesting that identification with a first-order desire still involves a higher-order volition in its favour, but is however not simply reducible to it. [32] In particular, the further element provided by Frankfurt to ground identification is the one of *deliberative decision*. Decision so understood represents a viable criterion for internality. This suggestion, Frankfurt says, is

[32] Cf. Bratman (2002)'s argument, above, 42.

immune to Watson's objection because "[d]ecisions, unlike desires and attitudes, do not seem susceptible… to externality" (Frankfurt 1975, 68).

At this point of the paper, then, Frankfurt proposes the following criterion for internality: the agent's desires are internal when they are brought about by the act of decision performed by the agent. Are decisions, however, really immune to externality? Velleman (1992) has claimed that such an appeal to decision does not work, because the agent can make "unwitting decisions" with which she does not identify. Velleman's example concerns a situation in which remarks of a friend provoke me to get angry and raise my voice. On later reflection, I recognise that the previous "grievances had crystallized in my mind. . . into a resolution to sever our friendship" (164). However, such a decision made by myself was unwitting, and now I understand that at the time of that decision I was not really identifying with it. As Velleman says, in a case like this "it was my resentment speaking, not I" (164-165).[33]

If Velleman is right, then it seems that a mere appeal to decisions is not enough to secure the agent's motivation from externality, as Frankfurt previously suggested. In Velleman's scenario, decisions seem at such susceptible both to internality and to externality. Moreover, Velleman's objection can be generalised, and one can claim that regarding any mental occurrence, even a decision, it is always possible to raise the question of whether or not the agent identifies with it. In this sense, identification cannot simply involve some actual mental act. On the contrary, for identification to work, it seems that one needs to appeal to a further grounding concept, which is not simply reducible to a mental act or occurrence.

2.4 Active Identification

In this section, I advance a reading of Frankfurt's claim (2), presented in the previous section.[34] I suggest that one way to make his proposal work, namely that decision is the element which prevents externality for higher-order

[33] Bratman (2002) notes that the unwitting decision of Velleman's example is merely a first-order one. Frankfurt might argue for this reason that Velleman is talking about choices and not decisions in the sense he endorses. However, Velleman's example could be easily rephrased so as to involve an unwitting decisions in favour of an higher-order desire, for example, the desire to severe the friendship. (6)

[34] See above, 48.

volitions, is by introducing a further grounding element for identification: *active control* of the agent with respect to her own desires. I call such a notion of identification *active identification (AI)*. However, I show that the outcome of such a reading is not acceptable in light of Frankfurt's overall theory of freedom.

I argue that decision works to justify identification if one appeals intuitively to the idea that freedom necessarily involves a voluntaristic act performed by the agent. Such a picture does not simply rely on the notion of decisive identification, but makes use of a further element to justify identification: the concept of *activity*. Frankfurt followed a similar strategy in his 1975 paper "Three concepts of a free action". In this paper, he tried to justify the role of identification by appealing to the notions of activity and self-determination. He writes:

> As for a person's second-order volitions themselves, it is impossible for him to be a passive bystander to them. They constitute his activity — i.e., his being active rather than passive — and the question of whether or not he identifies himself with them cannot arise.
>
> (Frankfurt 1975, 54)

In this passage, Frankfurt argues that the freedom of the second-order volitions is simply self-determining and as such not questionable, because higher-order volitions constitute the activity of the subject. In this picture, second-order volitions are free and exempt from externality because they are said to be – quite mysteriously - the active component in the agent's motivational structure. In this paper, Frankfurt adopts the activity of the second-order volitions as a notion which does not require a further explanation.

I suggest that in his 1987 paper, he complements this explanation through bringing the notion of *decision* into the picture. Indeed, the active element provided by Frankfurt to ground identification is deliberative decision. I argue that decision so understood can play the role of an act immune to externality because it introduces an element of *active control* of the agent with respect to her own desires. In this sense, the act of decision assures that the agent's desires are freely willed, because they are produced by the agent's own controlling

resolutions. In this light, freedom for higher-order volitions is secured because, with her deliberate decisions, the agent is able to 'create her own will,' thus assuring that she is free to will what she wants to will.

Such a notion of identification works not just by pointing to other higher-order phenomena in the agent's motivation, but instead is justified by the active decisions of the agent. I would like to call this notion of identification *active identification (AI)*:

> *(AI)* Freedom for higher-order volitions is secured when the agent identifies herself with the second-order volitions she decides to bring to alignment with her will.

In this picture, Frankfurt can justify his claims by appealing directly to the notion of activity: decisions are able to secure freedom because, when the agent decides in favour of a certain higher-order volition, she is active with respect to her will and not just a passive bystander to it. It follows that higher-order volitions brought about in such a way simply cannot be external to the agent's will.

In the rest of the section, I ask if such a strategy is viable, that is, if the notion of *(AI)* succeeds in providing a securing element for freedom and is itself immune to externality.

I suggest that Frankfurt can reply to this question by remarking that if one takes into account *(AI)*, then Watson's objections are simply misguided. Indeed, higher-order desires brought about by decisions simply *do not need* higher-order endorsement because the agent is active and not passive with regard to them. If the agent is not passive with respect to them, they simply cannot be external and, hence, unfree. As Cuyper (1998) notes, such a use of the notion of decision allows Frankfurt to escape Watson's objection because a decision is not something which is obscurely 'self-constitutive': decisions seem to be exempt from externality exactly because they are "act of the will" (47). Decisions are, in a sense, necessarily active: if the person identifies herself with her first-order desires by *making* a deliberate decision and producing the relevant intentions, it seems that such identification is an active rather than a passive process.

However, the appeal to the notion of activity in terms of 'deliberative decisions' and 'acts of the will' risks to provide only an insufficient explanation

to why, when endorsed by them, volitions are actually free. For which are the consequences of putting the burden of securing freedom on such kinds of act of the will? One unpleasant consequence is pointed out by Cuyper (1998, 48), when he notes that:

> Looked upon from this angle, Frankfurt's hierarchical model constitutes an attempt to reduce the notion of agent-causation within the explanatory framework of naturalism, yet in the end this project of reduction proves to be *circular*. He tries to account for autonomy and robust activity — that is to say, self-determination — in terms of reflective identification, but the explanation of decisive second-order volitions presupposes the primitive notion of agent-causation.

A charitable reading of Frankfurt suggests that he is ultimately not appealing to any form of agent-causation in order to justify the notion of decisive identification.[35] On this point, I agree with Velleman (1992) that Frankfurt's account of identification is meant to provide a reductionist account of agent-causation, in the sense that Frankfurt's aim is one of defining where *the agent* stands with respect to her desires.[36] However, I am additionally suggesting that the only way to make sense of his proposal, namely that decision is the element which secures freedom for higher-order volitions, is by introducing an active element in the agent: decisions have to be voluntaristic acts of the will and, as such, secure freedom.[37] Unfortunately, Frankfurt (1987) does not provide any further explanation about where to locate such a notion of activity.

In light of this, it is reasonable to conclude that Cuyper is right in claiming that the notion of activity, as it is used by Frankfurt, risks being in tension with the general naturalistic framework of his early account of freedom. Such a notion, if not incompatible with determinism, at least carries with it strong

[35] Frankfurt (1971) argues at length against Chisolm's account of free will as agent causation, and carefully distinguishes his proposal from these kinds of accounts.

[36] In this respect, Frankfurt's theory of action significantly differs from standard models of agency, which explain actions through a causal explanation. For a standard view of this kind, see Davidson (1980). Frankfurt argues against this model in his 1978 paper, "The Problem of Action".

[37] Frankfurt himself makes a similar remark in a later paper: "there is… a fundamental and ineradicable error in the very attempt to explicate being active in terms of endorsement or in terms of any other activity. Such attempts are manifestly bound to be circular" (Frankfurt 2002e, 205).

incompatibilist intuitions.[38] Since Frankfurt does not provide a more articulate story about decisive identification as a securing element for freedom, such a move could be a satisfactory reply to Watson's objection only if one accepts the persistence of incompatibilist intuitions of this sort.

2.5 Identification as Satisfaction

In this section, I examine the second amendment to the notion of identification presented in Frankfurt (1992). In this paper, Frankfurt proposes to explain identification with a further concept, the one of *wholehearted satisfaction*. My aim is to show that such an attempt fails, and that the role of identification as a ground for freedom is not clarified in an effective way in this later paper.

In his 1992 paper "The faintest passion", Frankfurt openly rejects his earlier attempt to understanding identification as a kind of decision, and he looks out for a different approach. Once again, Frankfurt starts his enquiry by considering the *presence of conflicts* in the motivational structure of an agent.

For Frankfurt (1971), conflicts in the agent's motivational structure signalled an absence of freedom for the agent. In that paper, Frankfurt suggested that when a conflict occurs, the agent is not exercising freedom of the will, because someone exercises free will only when she wants in complete accordance with pertinent higher-order volitions. In Frankfurt (1992), he explores a different notion of identification by emphasising a new dimension of such a motivational conflict: that of *ambivalence*.

For Frankfurt, someone is ambivalent when she is moved by incompatible preferences or attitudes regarding her affects, desires or other elements of her psychic life (Frankfurt 1992, 99). Accordingly, in Frankfurt's terminology ambivalence is not related to conflicts involving first-order desires alone (they

[38] Watson notes as well that Frankfurt stances toward compatibilism are ambiguous: "I should say Frankfurt's own attitude toward compatibilism is not exactly clear. There is a way of reading the condition that a person be 'free either to make that desire his will or to make some other first- order desire his will instead' in a way that is not captured by the dependency analysis, and he nowhere expliIcitly says that that analysis is sufficient as well as necessary for free will" (Watson 1987, 148).

"do not pertain to the will at all"). Instead, one is ambivalent with regards to one's higher-order, reflective attitudes. These kinds of struggles are "conflicts that pertain to the will": the will of an agent is affected by ambivalence when the agent is not decisively aware of what she really wants at an higher order of reflection.[39] In other words, with ambivalence the agent is left with a specific psychic situation that she is not capable to endorse or, instead, reject.[40]

In the spectrum of the possible psychological states of an agent, ambivalence is diametrically opposed to another state: *wholeheartedness*. Frankfurt says that an agent is "volitionally robust" (1992, 100) when she is resolved towards her higher-order attitudes, when she is, in other words, wholehearted in her desires, preferences and decisions. In this sense, a person is wholehearted when there are no conflicts or ambivalence among her higher-order volitions.

Wholeheartedness is clarified in Frankfurt (1992) through another notion. Indeed, according to Frankfurt, to be wholehearted is equivalent to the enjoyment of a kind of *self-satisfaction*. According to Frankfurt, when an agent is wholehearted with respect to some of her psychic elements, for example when she is wholehearted in some volition or attitude, she is at the same time "fully satisfied" with the circumstance that those motives determine her psychological processes. In this sense, to be fully satisfied with a particular psychological configuration of one's own motivational structure means that one has no active interest in changing it in any way. In other words, self-satisfaction is equivalent to a kind of absence of resistance with respect to one's motives and attitudes. In Frankfurt's words:

> Satisfaction with one's self requires no adoption of any cognitive, attitudinal, affective, or intentional stance … Satisfaction is a state of the entire psychic system – a state constituted just by the absence of any tendency or inclination to alter its condition… Being genuinely satisfied is not a matter, then, of choosing to leave things as they are or of making

[39] Note that a consequence of Frankfurt's notion of ambivalence is that an unwilling addict, who stands sincerely against his addiction, is not for this reason ambivalent with regard to his motivation. "The unwilling addict is wholeheartedly on one side of the conflict from which he suffers, and not at all on the other. The addiction may defeat his will, but does not as such disrupt its unity" (Frankfurt 2002, 99).

[40] In chapter 3, section 3.6, I will propose a different solution to the problem of ambivalence by making use of my hierarchical account of freedom.

> some judgment or decision concerning the desirability of change. It is a matter of simply having no interest in making changes.
>
> (Frankfurt 1992, 104–5)

The notion of self-satisfaction also allows Frankfurt to revise the notion of *identification*. Particularly, Frankfurt now characterises identification simply as the endorsement of a higher-order desire with which the person is satisfied (Frankfurt 1992, 105). In this light, if the endorsing desire is a desire with which the person is satisfied, she does not need to ascend to higher orders volitions to secure her identification with that desire.

2.6 Wholehearted Identification

In this section, I advance a reading of Frankfurt's notion of wholehearted satisfaction, that I presented in the last section. I propose to understand such a concept in terms of *wholehearted identification (WI),* and I show that *(WI)* does not work to ground freedom.

I argue that Frankfurt (1992) makes a decisive shift from an account in which identification is characterised as a form of *creation*, to one in which identification is better understood as a form of *discovery*. According to the first account, the agent's desires are her own desires because she actively endorses them by means of a higher-order volition or, alternatively, by means of an act of decision. According to this picture, identification is an act of the will. In this light, when one identifies with one's desires, "they are not merely desires he happens to have or to find within himself, but desires that he adopts or puts himself behind." In section 2.4, I proposed that such a notion of identification is well captured by what I called *active identification*. According to *(AI)* the freedom of lower-order desires is actively produced by higher-order volitions through the act of decisive identification. The higher-order volitions that the agent wants in alignment with her will are themselves free because they are established and produced in an active way through the act of making a deliberative decision.

In his later works, however, Frankfurt emphasises a different, more passive notion of identification. Frankfurt (1992) indeed suggests that this more

passive sense of identification is sufficient to ground freedom for higher-order volitions and to justify the role performed by identification. According to Frankfurt, identification is now not performed through an active decision, but it is performed instead in terms of satisfaction. Satisfaction with a higher-order desire, in turn, is merely a matter of an absence of conflicts in the motivational structure of the agent. In this light, identification is neither dependent on any form of production or 'creation' of the agent's will nor it does imply any form of action of the agent. On the contrary, identification is essentially a matter of passive discovery. In Frankfurt's words:

> To be satisfied with something does not require that a person have any particular belief about it, nor any particular feeling or attitude or intention. It does not require, for instance, that he regard it as satisfactory, or that he accede to it with approval, or that he intend to leave it as it stands. There is nothing that he needs to think, or to adopt, or to accept; it is not necessary for him to do anything at all.
>
> (Frankfurt 1992, 104)

In this picture, the agent identifies with her motives in a 'passive' way. Identification so understood does not presuppose any form of active or voluntaristic control by the agent. Nor it does presuppose a conscious acceptance of her volitional situation: the state of satisfaction with some psychic situation cannot be achieved in any active way. In what, then, does such a notion of identification consist?

On this account, one identifies with desires with which one is wholeheartedly satisfied. In this case, what the agent recognises is that she is satisfied with some of her physic elements, and when she recognises that this is the case, she automatically identifies herself with them. In this picture, most of the time the agent is not even aware of her satisfaction: she is simply identified with most of her psychic elements exactly because she is wholehearted in them.[41] I propose to understand such a different concept of

[41] With these remarks, I am not suggesting that the act of discovery is a necessary element for identification. Indeed, most of the time the agent is not even aware of her satisfaction with a certain desire. However, when the agent focuses on the scrutiny of her motives in the act of self-reflection, she would most of the time 'discover' that she is satisfied with most of her volitions and desires.

identification as a *wholehearted identification*. Accordingly, I formulate *(WI)* as follows:

> *(WI)* Freedom for higher-order volitions is secured when the agent is wholeheartedly satisfied with her identification with unopposed higher-order volitions.

(WI) is meant to sufficiently ground freedom for higher-order in quite a new way. In this picture, what secure the agent's freedom is not an arbitrary form of decisive identification with some of the agent's higher-order desires, but instead the fact that one's desires form a coherent and unitary set. *Wholehearted identification*, in this light, is fully realised in the act of recognition and acknowledgment of what one can wholeheartedly will. As Frankfurt notes, "we can only be what nature and life make us, and that is not so readily up to us" (Frankfurt 1992, 101). In the remainder of this section, I ask if the notion of *(WI)* succeeds in providing a grounding element for freedom and is itself immune to externality.

At first glance the notions of wholeheartedness and the one of satisfaction suggest a very different proposal than the one provided in Frankfurt (1987). Satisfaction, in particular, is not a further higher-order desire or attitude but, on the contrary, is a state of the entire psychic system of the agent. In this sense, being satisfied does not mean that the agent has another endorsing higher-order volition in her hierarchy. In contrast, when an agent is satisfied with a desire, she simply does not present any contrary attitude to it.[42] For this reason, since *(WI)* is not a further higher order desire, it seems a good candidate to solve the *Regress Problem*.[43] The main reason for this is that it removes the burden of securing freedom for higher-order volitions from any higher-order state as such. Instead, now Frankfurt chooses to understand the self not as a dialectic mesh between different levels of the person but as a single unity. This unity, of course, is still formed by different components, but in Frankfurt's new proposal what matters for freedom is the unity of the system itself. In this picture, the agent as a psychic system needs to be internally coherent in such a

[42] As Bratman (2002) eloquently puts it, Frankfurt's concept of satisfaction refers in a mere negative manner to a "non-occurrence" (7).

[43] Cf. chapter 1, section 1.10 for a detailed discussion of the *Regress Problem*.

way that, when any lower-order desire is presented to the agent, she does not have any contrary higher-order attitude toward it and she is instead wholeheartedly satisfied with it. When this happens, the higher-order volitions with which the agent is satisfied are as well free, exactly because of the absence in the agent of any tendency to reject them.

(WI) seems then promising in addressing the *Regress Problem*. Unfortunately, however, this concept raises two important issues. The first problem is that this notion itself seems to be not immune from externality. Indeed, it seems that under *(WI)* there is no clear way to distinguish between a free agent and a manipulated (or addicted) agent. This is a classical objection to hierarchical accounts of freedom, and it holds that it is possible that the alignment in a person's will could be the result of a manipulation, rather than being the result of the organisation of the agent's will. In other words, it is possible to imagine that the 'correct mesh' of will and volitions is implanted in a person from an external source (cf. for example Mele 1995).

I argue that such an objection is not met by Frankfurt's notion of *(WI)*. As I showed in the previous section, the understanding of satisfaction is fundamentally self-reflective ("the fact that the person is not moved to change things must derive from his understanding and evaluation of how things are with him", Frankfurt 1992, I05). However, if all there is to identification is self-reflection, self-reflection can indeed be susceptible to manipulation. In other words, it is possible to imagine that a psychic situation of wholehearted satisfaction is implanted by an external manipulator, but it is still perceived by the agent as free and autonomous. In this case, the agent is intuitively not free with regard to her satisfaction with her psychic situation: her feeling of wholeheartedness is not genuine, but it is instead externally implanted. However, the agent seems to meet Frankfurt's requirements for freedom: she is wholeheartedly satisfied with her psychological situation.

The reason for this failure, and this is the second worry, concerns the lack of a proper ground in Frankfurt's notion of *(WI)*. In Frankfurt's proposal, self-satisfaction is what grounds the notion of *(WI)*. Indeed, Frankfurt characterises identification simply as the endorsement of a higher-order desire with which the person is satisfied. In this light, if the endorsing desire is a desire with which the person is satisfied, she does not need to ascend to higher-order volitions to secure identification and freedom of the will. At the same time, since this kind

of identification is not a further higher-order phenomenon, it does not require a particular explanation: what in this picture assures that the agent's higher-order structure is free is simply an internally coherent set of physiological motivations. But is satisfaction enough to ground identification?

In the picture suggested by *(WI)*, the agent refuses to improve her situation even if she might know that she could be much better off, simply because she is satisfied with how things are with herself. Now, this situation allows two interpretations. On the one hand, the agent's acceptance can be motivated by the acknowledgment that her situation is "good enough" for her. On the other hand, however, it can simply be the result of depression or resignation. As Bratman (1996) points out, satisfaction with a desire "may be grounded in depression, and in such cases satisfaction with a desire does not seem to guarantee agential authority." (204) It seems that in this case the state of satisfaction with one's motivational situation does not express the will of the agent, but it is just the result of her passive state of depression. Satisfaction consequently, as the mere absence of contrary attitudes toward a desire, is not enough to firmly secure agential authority.

In light of this, it seems that the issue here is not to note that someone does not have any contrary attitude toward a desire, but is instead to ask *why s*he does not have any. In this respect Frankfurt's explanation of this notion seems incomplete. For why is an agent free if she is satisfied with certain motivations and not with others? Even if this process is entirely self-reflective and even passive, it still seems that we are left without a sense of why certain volitions with which the agent is wholeheartedly identified are meant to secure freedom for the agent.

Frankfurt could reply to this concern by pointing out that the agent is generally satisfied and identified with the volitions which form her *real self*. This further characterisation could help to grasp a better sense for which satisfaction with certain psychic attitudes guarantees freedom for the agent. However, as discussed previously,[44] what Frankfurt's argument seems to lack is a theory of personality, which is able to explain why some of the self's

[44] See chapter 1, section 1.10, 37.

volitions - the ones with which she is satisfied through *(WI)* – are supposed to form the real self.[45]

Alternatively, Frankfurt could reply by remarking that by being satisfied with the volitions which produce an alignment in the agent's will, the agent is not particularly identified with some volitions rather than others. On the contrary, the agent identifies herself with her entire psychic situation. In this light, the agent's will is free when there are not opposing tendencies in her system of motivations. But if this is the case, it seems that Bratman's objection still stands: the agent's state of satisfaction with her motivational situation could well be the result of a condition of depression or the like. In this sense, it seems that a mere negative characterization of the phenomenon of satisfaction is not adequate to justify a hierarchical theory of freedom.[46]

2.7 Summary of the chapter

In this chapter, I discussed Frankfurt's claim that identification is sufficient to secure freedom. I argued that Frankfurt presents two different notions of identification to support his claim, and I called them *Active Identification (AI)* and *Wholehearted Identification (WI)*.

I explored his 1987 paper, and I suggested that Frankfurt attempts to ground identification with the introduction of the concept of *decision*. Then, I investigated his 1992 paper, and I showed that here Frankfurt tries to secure the notion of wholehearted identification with the introduction of the concept of *satisfaction*. Through *(AI)* and *(WI)*, I presented a variety of arguments to understand whether Frankfurt's notion of identification could serve the

[45] In chapter 3, I present a revised hierarchical theory of free will which goes in a similar direction. Indeed, I propose a criterion to distinguish between desires that are, so to speak, part of the 'real self' of the agent and desire which are not, and are, as such, external. I will motivate these claims by proposing a different ground for the notion of identification: the one of *volitional identity*.

[46] On a general note, Frankfurt here faces the familiar objection that was discussed throughout the whole dissertation: he tries to justify freedom with a particular mental phenomenon, satisfaction, which is itself in need of an explanation. For this reason, other philosophers have tried to equip satisfaction with a different basis. For example, Bratman (1996, 2002) argues that one is satisfied with a desire *because of* a reason. For Bratman, practical reasons can be a proper ground for agential authority. However, in Bratman's theory what secures agential authority is the reason itself and not, instead, the same notion of satisfaction. In the next chapter, section 3.6, I will criticize Bratman's account with regard to the role of practical reason, and I will advance a proposal to incorporate practical reasons in a hierarchical account of freedom.

purpose of securing freedom. However, I concluded that both approaches are doomed to failure and that, as such, Frankfurt's notion of identification is not able to act as a ground for a hierarchical theory of freedom.

Ultimately, my analysis of the notions of *(AI)* and *(WI)* - combined with the distinction between three concepts of free will presented in chapter 1 - will serve as a basis for a new hierarchical account of freedom that I present in the next chapter.

3

The necessity of freedom

3.1 Outline of the chapter

In this chapter, I present a revised hierarchical account of free will. The main claim of my proposal is that freedom is dependent on the agent's volitional identity, which acts as a proper ground for freedom of the will. In section 3.2, I motivate why a hierarchical account of freedom should rest on a mere conformity account of free will *(Def. 3)* and on a notion of identification as satisfaction *(WI)*. Next, in section 3.3, I present three immediate objections to such an account. In the rest of the chapter, I articulate my account of freedom and provide a reply to these objections. As a first step, I propose a different notion of identification, which makes use of the notion of satisfaction, but it is grounded in the concepts of volitional identity and caring (section 3.4). Then, in section 3.5, I understand such a notion of identification as a ground for a hierarchical account of free will: free will, in this sense, is described as the agent's wholehearted identification with psychic elements which belong to her volitional identity. In section 3.6, I explore the problem of ambivalence in a hierarchical theory of freedom, and I suggest a way to overcome this motivational uncertainty according to my account. In section 3.7, I advance a proposal about the role of practical reason in a hierarchical theory of freedom. Finally, in section 3.8, I challenge my account by considering manipulation and addiction cases, and I propose an original solution to these problems.

3.2 A revised hierarchical theory of free will

In the previous chapter, I argued that Frankfurt makes use of two distinct concepts of identification to provide a ground for his theory of freedom. The general claim of Frankfurt's late proposal is that a correct account of freedom and autonomy rests on an adequate notion of identification. Indeed, to exercise free will is not just reducible to the correct mesh between agent's volitions and will, but instead it is a matter of *decisive identification with effective desires*, while a lack of such identification would entail the absence of free will. For Frankfurt, when the agent acts on a desire with which she has identified, she is acting as an active and autonomous agent. On the contrary, when she acts on desires with which she does not identify decisively, she is acting in a passive way, because she is moved by desires that she considers as alien to herself.

I suggested that Frankfurt makes use of two different notions of identification to explain how an agent is identified with her motives: *Active Identification (AI)* and *Wholehearted Identification (WI).* By making use of different higher-order phenomena - such as decision or satisfaction - these two concepts are aimed at justifying how the agent's higher-order desires can be in themselves exempt from externality and, hence, count as a proper element for grounding freedom. To repeat, according to *(AI)*, identification works as a securing element for freedom not just because it appeals to even higher levels in the hierarchy, but because it appeals to a decisive element of activity:

> *(AI)* Freedom for higher-order volitions is secured when the agent identifies herself with the second-order volitions she decides to bring to alignment with her will.

What *(AI)* suggests is that the freedom of lower-order desires is actively produced by second-order volitions through the act of making a deliberative decision. The act of decision, since it is actively performed by the agent, is not subject to any form of externality.

On the contrary, according to *(WI)* identification is simply the endorsement of a higher-order desire with which the person is satisfied. In this light, if the endorsing desire is a desire with which the person is satisfied, she does not need to ascend to even higher-order volitions to secure her identification with

that desire. I proposed to understand such a different concept of identification as a *wholehearted identification*. Accordingly, I formulated *(WI)* as follows:

> *(WI)* Freedom for higher-order volitions is secured when the agent is wholeheartedly satisfied with her identification with unopposed higher-order volitions.

According to *(WI)*, what secures the agent's freedom is not an arbitrary form of decisive identification with some of the agent's higher-order desires, but instead the fact that desires form a coherent and unitary set. *Wholehearted identification*, in this light, is fully realised in the act of discovery and recognition of what one can wholeheartedly will.

In this section, I ask how the notions of *(AI)* and *(WI)* work when applied to Frankfurt's (1971) original theory of free will. The aim of this move is to further investigate the role of the notion of identification in a hierarchical theory of free will. The result of this enquiry will show that a hierarchical theory of freedom correctly rests on the notion of identification, when such a concept is justified by the right grounding element.

In chapter one, I argued that Frankfurt (1971) does not present a unitary definition of free will, but instead is ambiguous between three different concepts of free will:

> *(Def. 1) Free will as counterfactual causation*. A person exercises free will if she is free to have the will she wants, and she could have constituted her will otherwise than she did.

> *(Def. 2) Free will as actual causation* A person exercises free will if she has the will she wants because it is the will she wants to have, and she has no interest in changing it.

> *(Def. 3) Free will as mere conformity*. A person exercises free will when she has the will she wants, even if she has no causal control over her will.

For the purposes of my argument, the difference between the three definitions of free will can be expressed by noticing that *(Def. 1)* and *(Def.2)*

suggest that there should be a *causal relation* between higher and lower motivational elements of the agent – higher level motivational elements should causally produce the lower level ones. On the contrary, *(Def. 3)* does not have to respect this condition: *(Def. 3)* suggests that to exercise free will higher-order volitions should be merely *in accord* with the will of an agent.

Now, as Watson's objection clearly shows, a hierarchical account of free will can work only if one is able to motivate which element in the agent's motivational structure is responsible for securing freedom for the higher-order volitions of the agent and, in turn, for the agent's will. The notion of identification seems very promising in this respect. However, I showed in chapter 2 that both the notions of identification put forward by Frankfurt are not exempt from structural problems, and, as such, they are not able to perform this task.

Frankfurt (1992) seems to sketch a possible reply to the question I am trying to answer, that is, what are the implications of the different notions of identification when applied to the concept of free will. Particularly, in this crucial passage, Frankfurt provides us with a sketch of how the shift from an account of decision to an account of wholeheartedness reflects on his theory of free will presented in his 1971 paper. Let me quote Frankfurt at length here:

> When the chips are down [a person] may discover that he is not, after all, decisively moved by the preference or motive he supposed he had adopted …We cannot have, simply for the asking, whatever will we want … This may appear in conflict with the notion that our wills are ultimately free. But what is the freedom of the will? A natural and useful way of understanding it is that a person's will is free to the extent that he has whatever will he wants. Now if this means that his will is free only if it is under his entirely unmediated voluntaristic control, then a free will can have no genuine reality; for reality entails resistance to such control. Must we, then, regard our wills either as unfree or as unreal? The dilemma can be avoided if we construe the freedom of someone's will as requiring, not that he originate or control what he wills, but that he be wholehearted in it. If there is no division within a person's will, it follows that *the will he has is the will he wants*. His wholeheartedness means exactly that there is in him no endogenous desire to be volitionally different than he is. Although he may be unable to create in himself a will other than the one he has, his will is free at least in the sense that he himself does not oppose or impede it.
>
> (Frankfurt 1992, p. 101-102, my emphasis)

To have a better understanding of Frankfurt's claims, I suggest looking back at the different concepts of free will at work in his 1971 article. In chapter one, I suggested that Frankfurt's (1971) theory of free will is equivocal between two general meanings of free will, which in turn give rise to the three different concepts of free will, restated above. According to my analysis, Frankfurt's original theory is ambiguous between two meanings of free will:

a) a person possesses free will if she is free to want what she wants to want.
b) a person possesses free will if she wants what she wants to want

In the quoted passage, Frankfurt notes that his previous attempts to characterize free will as *being free to have the will one wants*, that is, as meaning (a), are doomed to failure. The reason for this, Frankfurt says, is that to justify such an account and to reply to obvious objections to it one ultimately needs to resort to a form of voluntaristic and unmediated control. On the contrary, the weaker requirement of wholehearted satisfaction seems to suggest an alternative conception of freedom. In particular, Frankfurt is now advising favouring meaning (b) of free will, according to which free will is the freedom of simply *wanting what one wants to want*.

I consider Frankfurt's giving up on meaning (a) of free will a consequence of the failure of the notion of *(AI)*. I am going to show why. Let us consider the definition of free will which best suits meaning (a), that is *(Def. 1)*. According to it, a person exercises free will if she is free to have the will she wants, and she could have constituted her will otherwise than she did. To endorse *(Def. 1)*, it is indeed necessary to respect two different conditions:

(i) *Internal control*. In the actual scenario, the agent is free to have the will she wants. This means that there should be a particular relation between second-lever desires and first-level ones, namely that the second-level desires should causally produce the first-level ones.

(ii) *Counterfactual control*. Moreover, such a relation has to be confirmed as well in a counterfactual scenario. To be free, the agent "could have done

> otherwise than constitute the will as he did." That is, if the agent had wanted to want otherwise, she would have wanted otherwise.

By abandoning meaning (a) of free will, what Frankfurt is denying is that it is possible for an agent to actively exercise any form of internal or counterfactual control on her will. Such a kind of control, Frankfurt suggests, is not coherent with the reality outside us ("Indeed, the concept of reality entails the existence of something which is independent of our desires and wishes, and which, as such, constrains us", Frankfurt 1992, 100). A person, Frankfurt now says, cannot shape her will as she likes by a psychic movement that is fully under her immediate voluntary control (Frankfurt 1992, 1993, 1999b).

But how does Frankfurt reach such a conclusion? For the purposes of my argument, let us focus on the requirement of *internal control*. The latter condition suggests that an agent exercises free will if her second-level desires causally produce the first-level ones. As shown in chapter one, however, such a definition of free will is challenged by Watson-style objections. To repeat, Watson argues that, in order for such a definition to work, one has to explain *why* freedom for the agent is justified by the higher-level volitions' role. As shown in chapter 2, section 2.2, Frankfurt tries to find a justification for his claim in the notion of identification. In particular, adopting *(AI)*, Frankfurt suggests that the higher-level volitions guarantee freedom for the agent because she is able to decide, with regards to her higher-order volitions, which one will form her effective will. Decisions about higher-order volitions, Frankfurt suggests, are acts of the will, and as such are exempt from externality.

Internal control, in this picture, is then secured by the act of decisive identification of the agent with a higher-order volition, which is in turn justified by the agent's decision about it. However, I argued in chapter 2, section 2.4 that the only way to coherently make sense of this proposal is by introducing an active and voluntaristic element in the agent: only in this way decisions are able to secure freedom. What Frankfurt is denying in the quoted passage, however, is precisely that the agents' wills can be made free by any act of voluntaristic decision. As a matter of fact, Frankfurt now suggests that agents cannot simply originate or control their wills, and that free will has much more to do with the

agents being wholehearted in their wills and satisfied with their motivational structure.

It seems to me that the most natural way to interpret Frankfurt's claims is to say that to exercise free will, one does not need to be free to have the will one wants, as *(Def.1)* or *(Def.2)* suggest. On the contrary, it is sufficient that the will one has is the will one wants. Such a conception - as I argued in chapter 1, section 1.4.3 - is best captured by *(Def.3)* of free will, according to which a person exercises free will when her higher-order volitions simply align with her will, even if she could not have created the alignment or had another will instead. At the same time, Frankfurt suggests that meaning (b) of free will, that is, that a person possesses free will if she wants what she wants to want, is best suited to account for the weaker requirement of wholehearted satisfaction, that is, to the conception of identification that I called *(WI).*

In light of this, my aim in this chapter is to propose a revised hierarchical theory of free will by making use of *(Def. 3).* At the same time, I suggest that a coherent analysis of free will based on *(Def.3)* can rely only on a notion of identification as *Wholehearted Satisfaction*. *(Def. 3)* suggests that free will is better understood as a mere conformity between the higher-order volitions of an agent and her will: if they align, the person exercises free will. That is, if the person finds herself identified with aligned volitions, she is acting with a free will. So understood, *(Def. 3)* and *(WI)* jointly conjure to provide the following further definition of free will:

> *(Def. 4) Free will as self-satisfaction*: A person exercises free will if she is wholeheartedly satisfied with her identification with reflective elements of her psychic system, and she has no interest in changing them.

According to *(Def. 4),* a person exercises free will in a reflective and even passive way: the person's will is free exactly when she just finds herself wholeheartedly satisfied with the will she has, and she has no interest in changing her psychic situation. In this picture, free will is essentially a matter of complete and wholehearted satisfaction with regards to one's effective motivational structure.

3.3 Free will and satisfaction: an ostensible contradiction

The psychological picture suggested by *(Def. 4)* and, more generally, by the notion of satisfaction with one's motivational system, is in some respects quite realistic. Indeed, most of the time, agents show an attitude of satisfaction with the state of their motivational system: they act constantly on various desires and, if asked, they are confident in endorsing these desires at a higher-order of reflection.

To see how this happens, consider the following example. A few times per week, Susy loves to go out and eat junk food. When asked about her unhealthy habit, she is very confident about her reply: she loves the taste of it, and she thinks that to enjoy life and fulfil one's desires is more important than caring about one's health. She does this out of reflection and her desire to keep eating junk food is freely willed.

In terms of the account of free will suggested by *(Def. 4)*, Susy is indeed exercising free will. Susy has a first-order desire to eat junk food. At the same time, when she reflects on her motivational system, she finds in herself an alignment between her higher-order volitions and her first-order desires: she identifies herself with the second-order desire to keep desiring junk food and she is satisfied with her identification. As a matter of fact, she has no interest in changing her psychic situation, and she is wholeheartedly grounded in her volitions.

However, as discussed in chapter 2, section 2.6, the notion of satisfaction suffers from different problems, and it is not clear whether it is a proper candidate to justify identification. In the same way, it seems that, as a definition of free will based on the notion of satisfaction, *(Def. 4)* is immediately at odds with an acceptable conception of free will. Such an account, indeed, seems to raise at least three problems:

a. It conveys an idea of *passivity* which seems to conflict with an acceptable conception of free will.
b. It is not clear if satisfaction, as it is employed in *(Def.4)*, is sufficient to *ground* identification and free will.
c. It is not clear how to handle cases of *manipulation* and *addiction* under *(Def.4)*.

I begin by considering objection (a). According to this concern, Frankfurt's characterisation of satisfaction is at odds with free will since it seems to refer to a sort of *passivity* with regard to what moves one to action. In other words, the agent described by Frankfurt might know that she could improve her situation but she does not do anything to reach such an end because she is satisfied with the way things are. Now, the state of acceptance of one's motivational situation can derive from the appraisal that one's situation is 'good enough', and thus it does not require an intervention to improve it. However, as Bratman suggests, satisfaction can easily be the result of a state of depression or resignation. In this case, it is hard to see how satisfaction can express what the agent really wants, or as Bratman puts it, his authority as an agent. ("Why would that not suffice for identification? The answer is that one may leave things as they are because of some sort of enervation or exhaustion or depression or the like", Bratman 1996, 7)

In the example above, Susy endorses her first-order desire to eat junk food just because she is satisfied with it. However, this does not imply in any way an active commitment from her side to endorse a particular volition as the one she wants to be part of her effective will. Indeed, Susy's desire to eat junk food could easily be the result of a state of resignation or depression, and as such, could not express what Susy really wants.

Objection (b), as a consequence, states that satisfaction as the mere non-occurrence of negative attitudes toward a desire is insufficient *to ground identification and free will.* Since satisfaction can be the result of an altered psychic state such as depression or resignation, it seems that to claim that someone does not have any negative attitude toward a desire is insufficient to assure that her volitions are freely willed. For this reason, the real issue here is not that someone does not have any negative attitudes toward a desire, but rather *why s*he does not have any. In this respect Frankfurt's explanation of this notion seems incomplete. For why is an agent free if she is satisfied with certain motivations and not with others? Even if this process is entirely self-reflective and even passive, it still seems that we are left without a sense of why certain volitions with which the agent is wholeheartedly identified are meant to secure freedom for the agent.

Finally, objection (c) suggests a further unpalatable consequence of the agent's passivity under *(Def. 4)*. Namely, that it is not easy to distinguish between a free agent and an addicted or manipulated one. In the example, it is possible to imagine that Susy's second-order volition to keep desiring junk food is implanted from an external source, or that it can be the result of an addiction. In these cases, it seems difficult to claim that Susy is actually exercising free will.

Overall, it seems that a conception of free will based on Frankfurt's notion of satisfaction is strongly challenged by these objections. In the next sections, I will offer a reply to these objections by adding further elements as an appropriate ground for the notion of identification as satisfaction. My aim is to give at least some justifications for a theory of free will based on such a notion. In section 3.3, I reply to objection (b) by proposing an account of identification based on the notion of satisfaction, and grounded in the notions of volitional identity and caring. As a result, I propose the introduction of volitional identity as a new element to ground identification and free will. In sections 3.4, I argue that the structure of the will described in the previous section justifies a hierarchical account of free will based on the notion of satisfaction, and I present such an account. In section 3.5, I explore the problem of ambivalence and suggest a possible solution to it, and I offer at the same time an answer to objection (a). In section 3.6, I sketch a role for practical reason in my account. Finally, in section 3.7, I consider how my account of free will is able to deal with objection (c), that is, how it distinguishes between a free agent and an addicted or manipulated one.

3.4 Identification, caring and volitional identity

In this section, I argue that it is possible to make use of identification as satisfaction to build an adequate account of free will on *(Def.4)*, by taking into account Frankfurt's notions of *caring*, *love* and *volitional identity*. These phenomena, if appropriately connected, can provide the necessary ground to explain why most of the time agents do not have any negative attitudes toward their desires; moreover, they explain why the agent endorses at a higher-level

of reflection some of her desires and not others, and why her identification and satisfaction with them implies that her desires are freely willed.

In a nutshell, my claim is as follows: the agent's ability to care and love shapes her volitional identity. Volitional identity characterised as such grounds the notion of identification based on satisfaction *(WI)* as a securing element for freedom of the will. For this reason, satisfaction appropriately justifies *(Def. 4).*

I begin by analysing Frankfurt's notions of caring and love as he presents them in Frankfurt (1982, 1999). The concept of *caring* becomes particularly important in Frankfurt's later works. For Frankfurt, caring is the fundamental guiding principle of human actions. Furthermore, *love* is for Frankfurt the most important kind of caring and, as such, it plays an important role in understanding human agency. But what do caring and love exactly mean for Frankfurt?

Frankfurt (1982, 85) provides the first definition of caring:

> As for the notion of what a person cares about, it coincides in part with the notion of something with reference to which the person guides himself in what he does with his life and in his conduct … [Caring is] constituted by a complex set of cognitive, affective and volitional dispositions and states.

According to Frankfurt, caring is a composite disposition which guides the agent through her life, by means of providing her with the ability of "giving particular attention" (83) to certain things and directing her behaviour accordingly toward what she cares about. In the quoted passage, Frankfurt seems to claim that caring about something requires the presence of related emotions or beliefs. However, In Frankfurt (1993, 1999), he claims that it is not necessary to have any emotion or belief to care about something and that caring necessarily implies only a volitional element ("This is not primarily either a cognitive or an affective matter… Caring is essentially volitional; that is, it concerns one's will", 1993, 110).

Frankfurt (1999) explains what caring means by comparing it both with *wanting* something and finding something *valuable.*[47] An agent can have a strong desire for something, for example, the desire to eat junk food. However, to have the desire for something does not imply that an agent *cares* about it.

[47] See for example Frankfurt (1999b, 158).

Indeed, even if such a desire is strongly present in the motivational structure of the agent, she can care more about other things. For example, she can avoid eating junk food because she is worried about her health, and she cares about it. At the same time, it is also possible that an agent is willing to recognise that something is inherently *valuable*, for example doing exercise. However, such a thing is not something which the agent cares about. In this case, even if the agent is aware of the intrinsic value of exercising, she simply is not moved to act on such a consideration because she does not find it significant for her at all.[48] In this light, according to Frankfurt, to care about something means that something is in some respect *significant for the agent*, that she finds it important in guiding her conduct and actions.

Caring and wanting also differ in another significant respect. Indeed, one of the most important characteristics of caring is that it has a *temporal aspect* ("[T]he outlook of a person who cares about something is inherently prospective: that is, he necessarily considers himself as having a future", Frankfurt 1999, 83). It is possible for an agent to have desires and even beliefs without recognising them as elements integrated with her persisting psychological structure. In other words, having a desire or a belief does not entail that such a desire or belief will persist for a certain period of time. On the contrary, it is essential to the nature of caring that it persists for more than a single moment, and extends itself toward the future.

In this respect, caring also differs from satisfaction: according to Frankfurt, being satisfied with a desire means that the agent does not have any negative higher-order attitudes toward that desire in her motivational system. The role of caring, then, is one of extending satisfaction toward the temporally extended presence of a particular desire with which the agent is satisfied.

The ability of caring also contrasts significantly from voluntary *acts of will*, such as intentions, choices or decisions. Frankfurt (1982) recognises that the importance of such acts for the formation of a person's will is often excessive:

> This would hardly be worth pointing out except that an exaggerated significance is sometimes ascribed to decisions, as well as to choices and to

[48] One consequence of this distinction is that for Frankfurt value is in itself subjective. In his account, the value of the object that one loves and cares about derives from her caring about it. In this sense, it does not possess an objective value. Susan Wolf criticizes Frankfurt's concept of caring, and especially his subjectivist conception of value in (Wolf 2002, 227-244).

> other similar "acts of will." If we consider that a person's will is that by which he moves himself, then what he cares about is far more germane to the character of his will than the decisions or choices he makes. The latter may pertain to what he *intends* to be his will, but not necessarily to what his will truly *is*.
>
> (Frankfurt 1982, 84)

In the quoted passage, Frankfurt says that what persons care about is more central to their will than acts of the will as decisions or intentions, because what someone cares about delineates *what his will truly is*. As argued in chapter 2, section 2.3, Frankfurt had previously emphasised the importance of decisive commitment and decision when explaining identification and autonomy (Frankfurt 1971, 1975, 1977, 1987). Now, with the introduction of the concept of satisfaction, and then, with the notion of caring he claims, on the contrary, that the most important part of the will is something that agents cannot directly control or endorse:

> The formation of a person's will is most fundamentally a matter of his coming to care about certain things, and of his coming to care about some of them more than about others.
>
> (Frankfurt 1982, 91)

> Our essential natures as individuals are constituted, accordingly, by what we cannot help caring about. The necessities of love … mark our volitional limits, and thus they delineate our shapes as persons.
>
> (Frankfurt 1993, 138)

The role ascribed to caring by Frankfurt is thus quite fundamental. According to him, caring is an involuntary disposition which is proper to human persons. As such, caring guides the agent through her life, by means of providing her with the ability to give particular attention to certain things. The result of this process is a crucial one: the *formation of a person's will*. By caring about and loving different things over time, the agent's will is gradually and spontaneously shaped and, as such, it determines the effective activities of the agent in her volitional and agential life. The role of caring, then, is the

fundamental one of delineating the boundaries of a person's will, or, in other words, the limits of her volitional essence. But what does it mean, in Frankfurt's hierarchical terminology, to define the shape of a person's will?

For Frankfurt, a person is characterised by her ability to self-reflect about her motivational situation, and such an ability is realised in the formation of higher-order volitions. Such volitions, in turn, form the essential nature of a person:

> The essential nature of a person is constituted by his necessary *personal* characteristics … They are especially characteristics of his will ... The personal characteristics of someone's will are reflexive, or higher-order, volitional features.
>
> (Frankfurt 1993, 113)

Understood in this light, then, caring is not just simply a part of reflective self-evaluation of an agent or one more type of higher-order element which can justify the hierarchical organization of the will. On the contrary, caring is configured as the ability required *in order to* form higher-order desires and volitions. In other words, caring seems to be the ability which can *ground* the higher-order and self-reflective capacity of human persons. In this light, agents are able to form higher-order desires because they are able to care.

As a consequence, caring and love shape what Frankfurt considers the volitional identity of an agent. Frankfurt says that the volitional essence of a person is constituted by her higher-order, volitional features. As I am suggesting, however, such higher-order level of reflection is in turn grounded in the person's ability to care about different things. As a consequence, the essential nature, or volitional identity, of a person is accordingly constituted and shaped by what she "cannot help caring about", or, in other words, by things that come spontaneously to be important to her.

About certain things that are important to her, a person may care in such a way that she is subject to a kind of *necessity*. Indeed, Frankfurt says that there is a particular kind of necessity which is strictly linked to the configuration of

a person's will.[49] Frankfurt (1982), for example, presents the so-called Luther case. When Martin Luther was charged with heresy, he was offered an opportunity to recant by rejecting his heretical writings. However, he refused to do so, stating "Here I stand; I can do no other." In this case, the configuration of Luther's will prevents him from recanting. Even if he may intend to recant, he is subject to a volitional necessity, and he will inevitably fail to do so.[50] According to Frankfurt, Luther thereby discovers the boundaries of his will, which shape his volitional identity:

> When a person is subject to this sort of volitional necessity, it renders certain actions *unthinkable* for him. These actions are not genuinely among his options. He cannot perform any of them… he cannot *will* to perform them… If he attempts to do so, he runs up against the *limits of his will*.
>
> (Frankfurt 1993, p. 111)

The young man in Sartre's famous example is another instance of what a volitional necessity is. The young man resolves his dilemma, concerning whether to remain at home and look after his mother or to abandon her and join the fight against his country's enemies, by making a "radically free choice". However, according to Frankfurt (1982, 85)

> He might discover, when the chips are down, that he simply cannot bring himself to pursue the course of action upon which he has decided. Without changing his mind or forgetting anything, he might find either that he is moved irresistibly to pursue the other course of action instead or that he is similarly constrained at least to forbear from the course he has chosen ... That is, he might discover that he does not have and that he does not subsequently develop the feelings, attitudes, and interests constitutive of the sort of person which his decision has committed him to being.

The description of the young man's situation suggests that it is not sufficient to form a higher-order volition about how one wants his will to be

[49] The concept of volitional necessity is also discussed by Gary Watson. See Watson (2004), 88-122.

[50] Rober Kane claims that it is not true that Luther cannot change his mind about his beliefs. Indeed, Kane (1996) argues that Luther did have such an ability in his past. In this scenario, his moral responsibility and free will actually depends on Luther's (past) ability to act otherwise.

effective or even to commit oneself to a decision about it. Indeed, it could be that his higher-order volition is not integrated with other elements of his motivational structure, that is, he does not develop the appropriate attitudes, feelings or intentions to sustain and maintain his decision. In other words, that decision, based on some higher-order reflection, is clearly not part of the man's *volitional identity*: as such, it is impossible for him to fulfil his volitions *just because* he made a decision about his will. That decision is not between the things the man really cares about and, as such, he is not able to change is effective will accordingly.

3.5 Freedom as satisfaction with one's volitional identity

In this section, I argue that the structure of the will described in the previous section justifies an account of freedom of the will based on the notion of satisfaction *(Def. 4).*

It is possible to understand the young man's situation or Luther's situation in terms of Frankfurt's notion of satisfaction. In the case of the young man, he is not able to pursue his higher-order decision about his motives because he is not truly *satisfied* with the presence of that particular volition in his motivational identity. In Frankfurt's (1992) terminology, he is not wholehearted in that volition or decision and he is not satisfied with his will's configuration. Similarly, Luther is completely satisfied with the configuration of his will, and he is as such wholehearted in his volitions. Let us now recall *(Def. 4):*

> *(Def. 4) Free will as self-satisfaction*: A person exercises free will if she is wholeheartedly satisfied with her identification with reflective elements of her psychic system, and she has no interest in changing them.

According to *(Def.4),* a person exercises free will when she is wholeheartedly satisfied with her effective will, and she reflectively identifies herself with higher-order elements of her psychic system which are in line with her will.

I propose to analyse the situation of the young man and the one of Luther by making use of *(Def. 4)*. Accordingly, they reflect two opposite outcomes. Sartre's young man is not exercising free will according to *(Def.4),* because he is not wholeheartedly satisfied with his decision: he cannot bring himself to act on it, or he might find himself able to perform the action he has chosen, "but only by forcing himself to do so against powerful and persistent natural inclinations" (85). Luther, on the other hand, is exercising free will according to *(Def.4)*: the impossibility to recant is wholeheartedly endorsed by Luther, and he is satisfied with the configuration of his will *exactly because* he surrenders to a volitional necessity.

In light of this, what both cases suggest is that one exercises free will when he identifies himself with a higher-order volition, or with an element of his psychic system, if this element *is part of his volitional identity.* The fact that a higher-order volition belongs to the volitional identity of an agent justifies that the agent is satisfied with his identification with it and that he has no interest in making changes to his will. I propose to make use of Frankfurt's (1971) early theory of free will to explain how my proposal works.

At a lower level of volitional organisation, the agent possesses various and random desires. They derive from different factors, such as environmental stimuli, or the agent's biological constitution.

At a higher-level of reflection, the agent forms higher-order volitions in the act of reflection on her motivational situation. However, as argued in section 3.3, agents are able to form higher-order desires because they are able to care: in this sense, their will is shaped by what they come to care about during their life. More precisely, they form some of their second-order desires *in line* with what they care about. For this reason, some of the agents' higher-order volitions form their *volitional identity* .[51]

What *(Def. 4)* suggests, then, is that an agent usually identifies herself with first-order desires that are in line with the higher-order volitions which form

[51] I follow Frankfurt in claiming that *volitional identity* is essentially formed by higher-order desires and volitions ("The essential nature of a person is constituted by his necessary *personal* characteristics … They are especially characteristics of his will ... The personal characteristics of someone's will are reflexive, or higher-order, volitional features." Frankfurt 1993, 113). One consequence of this claim is that in my hierarchical account the higher-order volitions perform a fundamental role, the one of shaping the volitional identity of the agent. For this reason, even if in my account the higher-order volitions do not directly cause the lower-level ones, they are nonetheless essential in order to motivate freedom of the will.

her volitional identity: the agent exercises free will when she acts on desires with which she is wholeheartedly satisfied, and she is satisfied with them because they are part of her volitional identity. When she acts on first-order desires that are not aligned with her volitional identity, she does not exercise free will. By making use of the ground provided by the notion of volitional identity, *(Def.4)* can be further enhanced as follows:

> *(Def. 4.1) Free will as satisfaction with one's volitional identity*: A person exercises free will when she is satisfied with her identification with psychic elements which belong to her volitional identity, and she has no interest in changing them.

According to *(Def. 4.1),* then, freedom of the will is grounded in the volitional identity of an agent, which is in turn secured by her ability to wholeheartedly care and love. [52]

In the rest of the section, I consider one potential complication for my proposal and I motivate how my account replies to objection (b), presented in section 3.3.[53]

One possible problem for my account lies in the possibility for agents of acting upon various psychological elements, which are not immediately related to higher-order volitions. For example, let us consider again the case of Sartre's young man and his dilemma, concerning whether to remain at home and look after his mother or to abandon her and join the fight against his country's enemies. What if the young man does not leave his home out of a different psychological element, for example, fear? It seems that in this scenario his decision is not taken according to his volitions, but out of a complete different emotion. Does his action count as free according to my account?

I suggest that here two possibilities stand. One the one hand, it is possible that the fear and the young man's volitional identity are not in line. In this scenario, the young man's higher-order volitions are in line with the will to leave. However, because of fear, the man decides not to go. In this case, the

[52] In his (1982), Frankfurt claims something similar when he writes that a "person acts autonomously only when his volitions derive from the essential character of his will" (132). However, he does not provide any argument or explanation to motivate this intuition.

[53] See above, 68-69.

action does not count as free, because the fear prevents him from acting on volitions which are part of his volitional identity. On the other hand, it is possible that the fear and the volitional identity go in the same direction. In this scenario, the man does not leave out of fear either. However, upon reflection, he discovers that he does not have a higher-order volition to leave. In this case, his action of not leaving counts as free, because he is acting in line with higher-order volitions which are part of his volitional identity.[54]

I consider now a more general concern. I am going to ask if my account is adequate to reply to objection (b) (section 3.3.), that is, to provide a suitable ground for freedom of the will. Let us recall the objection briefly. According to objection (b), satisfaction as the mere non-occurrence of negative attitudes toward a desire is insufficient to ground identification and free will. Indeed, following Watson-style objections, one can ask *why* the fact that an agent is wholeheartedly satisfied with some higher-order desires is meant to secure freedom for the agent.[55]

In a nutshell, I suggest that identification as satisfaction can serve as the securing element for freedom of the will if it is grounded in the concept of volitional identity *(Def. 4.1)*. In order to further motivate my claim, let me briefly recall why it is important to find a grounding element to motivate an account of freedom of the will. According to the understanding of the free will problem that I advanced throughout this dissertation, the main question about freedom can be described as a question about which elements ground the ability to exercise free will.[56] In this light, then, to ask what grounds freedom of the will means to ask what circumstance or condition can make the case that the agent is exercising freedom. In other words, when we ask if an agent is choosing freely, what we want to find is one or more conditions *in virtue of which* her choice is a free one. My analysis suggests that what is missing in Frankfurt's account is exactly a proper ground for the notion of satisfaction

[54] On a general note, it is possible to argue that an element of uncertainty remains, because it is not always possible to easily distinguish between volitions and other emotions which motivate an action. However, my answer still stands, as it is not the fear - or any other emotion - in itself which determines whether an action is or is not free.

[55] See above section 3.2, 69.

[56] I understand *grounding* in the sense suggested by Sartorio (2016a, 8): "'grounding' [is a] placeholder for some relation of non-causal, ontological dependence between facts in light of which it makes sense to say that certain facts are more basic or more fundamental than others, in that those other facts obtain because of, or in virtue of, them".

and, in turn, for freedom.[57] My aim, then, is to argue that volitional identity provides such a ground, and help explain why an agent exercises free will when she is satisfied with her motivational situation.

To see if this is the case, I ask if my proposal is as well affected by Watson-style objections, as it is arguably the case with Frankfurt's original notion of satisfaction. I am going to show that the answer is negative. According to my suggestion, the agent's exercising a free will when she is wholeheartedly identified with her desires is not secured by the mere fact that she endorses a particular higher-order desire with a decisive commitment, or by her satisfaction with her higher-order volitions. Such answers, as showed throughout the dissertation, are not suitable replies to Watson's objection because they merely push the objection one step further.

On the contrary, according to my proposal, an agent is free when she is identified with a desire *because* such a desire is part of her volitional identity, which is in turn shaped by her higher-order reflection about what she cares and loves. The agent's desires can ground freedom of the will not simply because they are higher in the hierarchical structure of the will, or because the agent is satisfied with them, but because they are 'essential' to the subject's volitional identity. Such a volitional identity is so powerful to impede the agent from identifying with desires which are not part of her volitional identity. For the agent to identify with some desires and not with others is, indeed, a kind of necessity, a volitional necessity. In the picture I am suggesting, the agent is usually satisfied only with desires that are part of her volitional identity. In this light, volitional identity allows one to avoid any risk of arbitrariness when motivating the agent's identification with some of her desires. In the rest of the chapter, I will further motivate my claims.

[57] See also Bratman (1996) for a similar concern about the lack of grounding in Frankfurt's theory of freedom.

3.6 The problem of ambivalence

In this section, I explore the problem of ambivalence in a hierarchical theory of freedom, and I suggest a way to overcome this motivational uncertainty according to my account.

Frankfurt (1992) characterizes ambivalence as a kind of "disease" which paralyses the motivational structure of an agent:

> Insofar as someone is ambivalent, he is moved by incompatible preferences or attitudes regarding his affects or his desires or regarding other elements of his psychic life. This volitional division keeps him from settling upon or from tolerating any coherent affective or motivational identity. It means that he does not know what he really wants.
>
> (Frankfurt 1992, 99)

What this passage suggests is that ambivalence only arises out of a person's higher-order, reflective attitudes. On the one hand, the reflective volitions are manifestly opposed in the agent's psychological situation; on the other hand, they are both internal to a person's will and not "alien" to him. For this reasons, the presence of a conflict in the higher-order level of the person prevents her from forming a motivational identity with which she is satisfied.

It is a matter of fact that often an agent can be unsatisfied with the configuration of her will or be uncertain about what she really wants at an higher-level of reflection. If this happens, there are two possibilities. First, the agent can act on her first-order desire and, in this case, according to *(Def. 4.1)* her will is not free because there is, in her motivational system, a higher-order volition which is opposed to that first-order desire. Secondly, however, the agent can try to change the configuration of her effective will by identifying herself, at a higher-level of reflection, with an opposed higher-order volition. The aim of the process would be one of securing that the agent's desires are freely willed and are aligned again, or that, in Frankfurt's words, the agent wants what she wants to want.[58]

[58] In Frankfurt's perspective at that time, the presence of this ability is fundamental, because it is necessary for distinguishing between an addicted, or manipulated agent, who cannot bring in line her motivations, and a free agent, who is always able to do so. According to Stump (1988), for Frankfurt an agent exercises free will only if she has the first-order volitions she has

Frankfurt's explanation of this phenomenon in 1971 involved assuming that there should be a dependency relation between higher-level desires and lower-level ones: if the agent wants, she can bring in alignment her first-order desires by making a decisive commitment toward one of her second-order volitions. In terms of my definitions of free will, the will of an agent is free if she is able to causally produce the alignment via internal control (*Def.1* or *Def.2*). To explain this process, and to reply to Watson's objections, Frankfurt makes use of different higher-order phenomena, such as decisive identification or decision and commitment. However, as I argued in chapter 2, these attempts are doomed to failure and Frankfurt himself now rejects such a view.[59]

How, then, is it coherently possible for the agent to overcome ambivalence in an account of the will based on the notion of satisfaction? According to Frankfurt (1992, 100):

> Since ambivalence is not a cognitive deficiency, … it cannot be overcome voluntaristically. A person cannot make himself volitionally determinate and thereby create a truth where there was none before, merely by an "act of will." In other words, he cannot make himself wholehearted just by a psychic movement that is fully under his immediate voluntary control.

However, the absence of any effort on the agent's part to make herself wholehearted is problematic. Indeed, according to this picture, Frankfurt's notion of satisfaction seems to be too passive, and not able to account for changes in the will of an agent which result in the possession of a free will. Yet, it seems that if an agent is ambivalent, she can be able to make up her mind about how she wants to be motivated and to regain freedom of the will again. For these reasons, satisfaction, as presented by Frankfurt, seems to be even more implausible as a securing element for free will.

I suggest that a possible reason for this failure is to be found in the *incorrect relation of dependency* that Frankfurt presupposes between volitions of a different order. I propose to re-define such a relation by considering the effect of the external world on the motivational system of the agents.

---

because of her second-order volitions, that is, her second-order volitions have, directly or indirectly, produced his first-order volitions.

[59] Cfr Frankfurt (1992, 1999, 2004).

Let us recall the volitional situation of Susy in the scenario of section 3.2. Susy possesses a first order desire to eat junk food. However, she is now ambivalent at a higher-level of reflection about whether she wants to be motivated by such a desire. In other words, she possesses two opposite and conflicting higher-order volitions: she wants to keep desiring junk food because she loves and enjoys it; at the same time, she wants to desire to eat healthier foods because she is seriously concerned about her health. Now, Susy is able in the end to take a step to resolve the ambivalence by favouring the higher-order volition of desiring healthy food. How is it possible for such a volition to become wholehearted and hence effective in moving Susy to act *without any voluntaristic effort* on the agent's part, if such a volition is not at the moment part of the volitional identity of the agent and she is not fully satisfied with it?

I propose to understand the role of higher-order volitions in this picture as *directing our attention* to some features of the external word which are coherent with that second-order volition. In the case of Susy, for example, her higher-order volition to desire healthy food directs her attention in the external world to circumstances related to her volition: she reads about the relation between health and food, she buys health magazines, or she talks to people about such matters. As a consequence, it may happen that her first-order desires *change* in accordance with that second-order volition. In the example considered, Susy can, in the end, develop a first-order desire to eat healthy food as a result of influences from the environment and, in so doing, she regains an alignment between her higher-order volitions and her first-order desires.

Crucially, in the picture I am suggesting the causal relation that motivates changes in the will of the agent has not to be found *among volitions of different orders*, as in Frankfurt's proposal. On the contrary, I argue that such a causal relation exists *between the external world and the will of an agent*. Because of this relation, it can happen that the first-order desires of an agent change in accordance with her second-order volitions, without presupposing a condition of internal control between the different orders of volitions.

It can be useful to dwell a little more into the role that higher-order volitions play in my proposal. I am suggesting that the main role of second-order volitions is not that of securing free will by causing the alignment in the will of an agent, but instead that of directing her attention toward different features of the external world. As I argued in section 3.3, however, the ability

to form second-order volition is grounded in the agents' ability to care and love. In this sense, what guides our attention to the external world is as well grounded in our disposition to care. Frankfurt seems to suggest something similar when he writes:

> His caring about it consists, rather, in the fact that he *guides* himself by reference to it. This entails that he purposefully direct his attention, attitudes, and behavior in response to circumstances germane to the fortunes of the object about which he cares.
>
> (Frankfurt 1993, 110-11)

However, what is missing in Frankfurt's proposal is an explanation of how a fully-fledged notion of freedom can be guaranteed according to a satisfaction model.

My proposal, in this respect, seems to be more promising in explaining how changes in caring shape the agent's will and, at the same time, ground the agent's freedom. Indeed, an important feature of my proposal is that it does not presuppose any "deliberate volitional control" from the agents to change the configuration of their will. What an agent needs is simply to be wholeheartedly satisfied, according to *(Def. 4.1)*, with the new configuration of her will. When she is, she possesses a free will.

At the same time, however, the agent is not a "passive bystander" to her own volitional situation: by means of favouring an higher-order volition or simply entertaining it in her mind, the agent plays an active role in influencing the effective configuration of her will. What happens next, however, is just the result of the agent's spontaneous disposition to care for certain things instead of others.

3.7 A role for practical reason

In this section, I advance a proposal about the role of practical reason in a hierarchical theory of freedom. The role of practical reason in agency is the subject of an extensive debate, which is naturally interrelated with hierarchical theories of motivation. In his late works, Frankfurt started taking this literature

into account, trying to carefully situate his position in the debate (See for example Frankfurt 1999b, and, especially, 2004). I cannot hope to deal with this subject satisfactorily in this dissertation. What I offer in this section, however, is the starting point of a line of argument which shows that reasons can have a place in a hierarchical theory of freedom.

In the last section, I proposed to understand higher-order volitions as *directing our attention* to some features of the external word which are coherent with those second-order volitions. But in which way do higher-order volitions direct the agent's attention toward the external world? My tentative hypothesis is that what we care about, in terms of higher-order volitions, directs our attention to the external world by *providing practical reasons* upon which agents reflect. My claim, in a nutshell, is that practical reasons are the effective tool through which higher-order desires contribute to shaping the agent's will. In the remainder of the section, I will try to provide a sketch of such proposal.

I begin by asking what is the role of practical reasoning. One plausible answer to this question is that agents make use of practical reason, through deliberation, to make up their mind about how to reach a particular goal that they desire to reach. In so doing, the agent is lead to assess her motivational situation, by reflecting on her own motives and reasons in a practical way. In this respect, practical reason helps to figure out what is the best way to reach a specific end. It is a fundamental aspect of Frankfurt's proposal, however, that such a picture of practical reason, if not incorrect, is at least incomplete. Practical reason alone is not able to fulfil this role. Indeed, according to Frankfurt, to explain the agent's ability to reflectively evaluate her situation, one has to take into account the effect of the volitional level of the agent, that is, what the agent *desires* or *cares about* the most.

The former view of practical reason is shared, for example, by Michael Bratman. According to Bratman, even if identification, as Frankfurt understands it, needs not involve a commitment to forms of practical reasoning, it is plausible to think that the appeal to it is fundamental in explaining our agency. In this sense, there should be in the agent's motivational system a phenomenon that is able to account for the role of practical reason. Let us look at Bratman's solution to this problem. Bratman proposes to interpret Frankfurt's claim that "the agent identifies with first-order desire D

as functioning as an effective motive" in a way that takes into account practical reason and deliberation:

> (1) The agent endorses her treating D as providing a justifying reason for action in motivationally effective practical reasoning.

In this sense, a person is identified with a desire D if: (*i*) she treats it as reason-giving, (*ii*) she does not treat it as external, and (*iii*) she would decide to continue to treat it as reason-giving, be satisfied with that decision, and continue to treat it as reason-giving if she were to reflect on the matter (cf. Bratman 2002).

Bratman emphasises that satisfaction with a desire is a crucial element of human agency. However, he thinks that this concept is not enough to explain agency, if it is not paired with that of practical reason. In this sense, he claims that one is not simply satisfied with a desire but with treating it as reason-giving and continuing to treat it that way. That is, satisfaction is justified and grounded in a further phenomenon: that of treating a desire as a reason.

One obvious objection to Bratman's account is that he seems to treat the concept of deciding to treat as reason-giving as a primary concept, in the spirit of Frankfurt's (1987) proposal about decision. In this sense, it is not clear if the introduction of a further phenomenon such as the one of "deciding to treat a desire as reason-giving" could count as a sufficient ground for the notion of identification, or if it is, instead, subject to the same kind of objections that have been raised against Frankfurt's notion of decision.[60] Here, I will not attempt a detailed discussion of Bratman's account. Nonetheless, as a general remark, I would like to emphasise that the overall spirit of Bratman's proposal is very distant from Frankfurt's conception of the will and its relation to practical reason.

Frankfurt decisively rejects Bratman's normative interpretation of his claim that "the agent endorses a desire as a legitimate candidate for satisfaction." For Frankfurt, the higher-order attitudes that are formed in processes leading to identification "involve 'evaluations' only in a sense that is

[60] Bratman himself is aware of such a risk, and in 1996 he tries to argue why this is not the case. ("But what does it mean to treat a desire as reason-giving? Is it, in short, to identify with that desire? Are we moving in a circle?" (9). Also, cf. chapter 2, section 2.3 and 2.4 for a discussion of the objections to Frankfurt's notion of identification.

strictly value neutral" (Frankfurt 2002, 87). Frankfurt (1999) seems to sketch a possible link between the volitional level of the agent and her practical reasoning which goes in a very different direction. Indeed, he suggests that:

> The immediacy of the linkage between loving and what counts as a reason for doing things that help the beloved is part of what *essentially constitutes* loving. A person will not take the fact that a certain action would fulfill a duty as a reason for performing that action *unless the person has a desire* to do what duty demands… His taking it as a reason for performing the action is not the outcome of an inference… His taking it as a reason is a constitutive aspect of his loving: to love a person *is* essentially (in part) to take the fact that a certain action would be helpful to that person as a reason for performing it.
>
> (Frankfurt 1999b, 176)

Frankfurt's interesting suggestion here is that an essential component of the dispositions of caring and loving is that of providing the agent who cares about something with *reasons* to care. Therefore, the effect of practical reasons on the agent's motivational situation depends on the fact that such reasons essentially *desires* of the agent, because they are provided by her spontaneous disposition to care about things. Otherwise, the agent would not be interested in that reason and such a reason would not be effective on her will.[61]

I argue that such a suggestion is not taken into account by a proposal such as Bratman's, which considers the phenomenon of deciding to treat a desire as a reason as the ultimate ground for identification. On the contrary, I propose a role for practical reason that is more in line with the spirit of Frankfurt's proposal. I am arguing that what we care about, in terms of higher-order volitions, directs our attention to the external world. In doing this, our volitional disposition to care provides us, at the same time, with practical reasons upon which agents reflect. In other words, higher-order volitions provide reasons for agents to 'make up their mind' and reflect about their motivational system.

However, this does not imply that the agent exercises any act of decision about her desires, or that practical reasons, as a separate element in the agent's

[61] For further criticisms of Frankfurt's conception of reason see also Scanlon (2002) and Moran (2002).

motivation, have a causal efficacy on the agent's will. On the contrary, in my proposal they fulfil their role exactly because they are, at the same time, the expression of desires of the agent: what the agent cares about is essentially presented to the agent herself in the form of practical reasons for caring about something in a certain way. In this light, even if reasons do not fulfil any separate role with respect to the agent's will, their role is quite fundamental: they are a helpful tool for the agent to participate in the process of deliberation and to shape her effective will.

3.8 Manipulated and addicted agents

In this section, I take into account objection (c) (cf. section 3.2), which says that it is not clear how to handle cases of *manipulation* and *addiction* under *(Def.4.1)*.

Let us first consider the problem of manipulation. The manipulation objection is a classical worry about hierarchical theories of free will. In its general form, it says that it is possible that the alignment in a person's will could be the result of a manipulation, rather than being the result of the organisation of the agent's will. In other words, it is possible to imagine that the 'correct mesh' of will and volitions, which is responsible for free will, could be implanted in a person from an external source.[62]

Consider the example, presented by Mele (1995), of two agents, Ann and Beth. Until a certain moment, they are two different persons, with different psychological structures. After a certain time, however, a neuroscientist implants in Beth a psychological structure which is identical to the one that Ann has acquired on her own. At such a point, there is no difference as regards to the structure of the wills of the two agents. But if this is conceivable, then if Ann acts on her own free will, Beth, who acts exactly in the same way, acts in a freely willed way too. However, such a result seems strongly counter-intuitive: if Beth's will is the result of a neurological manipulation, how can the actions that issue from her will be freely willed?

[62] For different perspectives on the manipulation-objection see Fischer (1994), Kane, (2002b), Pereboom (2001) and van Inwagen (1983).

I argue that *(Def. 4.1)* is able to distinguish between the two cases and help us in dealing with problems of manipulation. *(Def. 4.1)* suggests that a person's will is free when she is satisfied with her identification with elements of her psychic system which belong to her volitional identity. In other words, the volitional identity marks a criterion for internality for the agent's desires: when they are part of the agent's volitional identity, they are internal to the agent's will. Accordingly, *(Def.4.1)* suggests that in addition to the agent's being satisfied with a desire, the latter must also originate from the agent's volitional identity for her to identify with it and exercising free will.

By looking back at Mele's example, it is clear that Ann's desires are part of her volitional identity: as such, they are good candidates to be considered freely willed. On the contrary, one can wonder if Beth's desires, which have been implanted in her will by a neuroscientist, are as well part of her volitional identity.

It seems to me that the way in which the volitional identity of an agent is formed strongly discourages the interpretation of Beth's desires as part of her volitional essence. As I argued in section 3.4, the volitional identity of an agent is gradually shaped by her spontaneous disposition to care about things. By caring and loving over time about different things, indeed, the boundaries of the agent's will are gradually and spontaneously formed and, as such, they determine in that particular moment which desires are part of the volitional identity of the agent (and are, as such, internal to her will), and which in contrast are not (and are, as such, external and alien to the agent's volitional structure).

A first reply to Mele's case, then, can be that Beth's desires are not free because they are not formed *over time* by her natural disposition to care about things. On the contrary, indeed, they are implanted in her motivational structure by an external intervention, namely, the one of the evil neurosurgeon. Accordingly, Beth's desires seem to be non-free under *(Def. 4.1)*, which states that a desire is free if it originates from the agent's volitional identity. However, imagine now that the evil neurosurgeon does not manipulate Beth's will in a single time, but instead that he implants Ann's psychological structure in Beth slowly and over time, until the motivational situations of the two agents perfectly coincide. We may ask: in such a scenario, are Beth's desires freely willed?

To answer this question, I suggest looking at yet another feature of volitional identity, namely that it is usually formed *spontaneously* from the agent's natural disposition to care about different things. It seems to me that the desires which derive from a manipulation are not part of the normal psychological development of the agent's will in the spontaneous way just described and, as such, they do not satisfy this criterion. For this reason, Beth's desires, even if actually identical to Ann's desires, are not part of her volitional identity: they are not in this respect internal to her will and consequently they are not freely willed by the agent.

Let us now consider cases of addiction, by distinguishing between an unwilling addict and a willing addict. *(Def. 4.1)* can easily account for the distinction between an unwilling addict and a person who exercises free will. The unwilling addict presents conflicts at a higher level with what she wishes to be: her action of taking the drug does not issue from a second-order volition which is aligned with her will. The unwilling addict, then, is not able to exercise free will because there is a discrepancy between her will and her second-order volitions. In terms of *(Def. 4.1),* therefore, the unwilling addict is not *satisfied* with some elements of her psychic system (namely, her desire to take the drug), and she is not exercising freedom of the will.

On the contrary, the condition of satisfaction with one's own volitions closely resembles the status of a willing addict. The willing addict has conflicting first-order desires as regards taking a drug. But the second-order volitions of the willing addict are in line with her addictive first-order desire to take the drug: she wants to take the drug and wants her will to be formed by such a desire. Accordingly, under *(Def. 4.1)* it seems *prima facie* difficult to distinguish between the willing addict and a person who acts on her own free will, because the willing addict appears to be wholeheartedly satisfied with her motivational situation. The problem is, however, more general: how is it possible to distinguish under *(Def.4.1)* between a desire which is part of the volitional identity of the agent and one which is instead generated by an addiction?

The issue is very controversial. However, I think that the best answer to this objection for a hierarchical theorist would just be to bite the bullet. I think that it is indeed structurally impossible to fully distinguish between addicted

desires and free ones.[63] Such a distinction is not a clear-cut one, and it is in line with my proposal to claim that this is actually the case. One can note that the difference between addictive desires and free ones comes in degrees, and in some cases (as in the case of Susy's desire for junk food) it is intrinsically difficult from a psychological point of view to judge if a desire is a free or an addictive one, and, as a consequence, if it is really part of the volitional identity of the agent. The consequence of these remarks, I take it, is that the willing addict, under *(Def.4.1),* is actually exercising free will.

Such a partial result, however, presents at least an advantage with respect to Frankfurt's original theory of free will. I am going to explain why. In chapter 1, section 1.5, I argued that the willing addict does not exercise free will under Frankfurt's conditions. By looking back at the different definitions of free will found in Frankfurt (1971), Frankfurt seems to think that the willing addict exercises freedom of the will according to *(Def. 2).* On *(Def. 2)*, the willing addict is free if she satisfies *actual causation*, that is, her higher-level motivational elements should be in a direct, causal relation with her lower-level ones. At first glance, it seems that in the case of the willing addict the causal relation between higher-level desires and lower-level desires is not respected, because the willing addict's desire is effective due to her physiological addiction. In this light, if the willing addict's first-order desire for the drug is due to an addiction, it is not correct to say that her higher-level desires causally produced her lower-level ones.

However, Frankfurt seems to claim the opposite, that is, that when the willing addict takes the drug, she "takes it freely and of her own free will." Accordingly, the only way to say that the willing addict exercises free will is to claim that the willing addict's second-level wants *do* causally produce the first-level ones. However, to accept this conclusion one has to postulate, as Frankfurt does, an *overdetermination* of the willing addict's first-order desires to take the drug: her desire is effective because she is addicted to the drug but, at the same time, it is her effective desire because she wants it to be (Frankfurt 1971, 25).

[63] Cf. on this issue Dill & Holton (2014). In this paper, the authors argue that addictive desires are not substantially different from non-addictive desires. On their account, what distinguishes between them is the way in which they are acquired, which in turn regulate their strength. For the authors, one consequence of this account is that the effort to achieve self-control for addicted agents is not different in kind from the one needed by non-addicted agents. Rather, it differs in degree and difficulty.

If, however, one wants to reject the overdetermination strategy, one has to conclude, in contrast with Frankfurt's position, that the willing addict fails to exercise free will even according to *(Def. 2)*. By contrast, I showed above that the willing addict exercises free will under *(Def.4.1)* without the need of invoking overdetermination, because she is satisfied with her addictive desire to take the drug.

The only way I can think of to show that the willing addict is not exercising freedom and that her situation is in a sense different from the one of a free agent is to appeal to a *ceteris paribus* condition. That is, one can claim that an addiction, as such, is an element which interrupts the normal psychological development of an agent. In this respect, the case of the willing addict resembles the one of a manipulated agent. In the psychological development of the willing addict, there is a physical impediment to make her non-addicted higher-level desires effective in contributing to shaping his will. In this sense, one can claim that the willing addict lacks a proper volitional structure to secure freedom of the will.

3.9 Summary of the chapter

In this chapter, I presented an original hierarchical account of free will. Building upon my analysis of Frankfurt's account of freedom developed in chapter 1 and 2, I presented an account which combined the main elements of Frankfurt's original theory (*hierarchical levels of desires, identification*) with new elements from Frankfurt's late reflection about the psychological structure of agents (*caring, volitional identity*). Accordingly, I advanced a detailed account of the psychological structure of agents, and I argued that such a rich structure of the will is what grounds my hierarchical account of freedom.

The main aim of my proposal was to equip a hierarchical theory of free will with a different and effective grounding element, the *volitional identity* of the agents. In a nutshell, I suggested that identification as satisfaction can serve as the securing element for freedom of the will if it is grounded in the concepts of volitional identity and caring. Accordingly, I understood free will as the agent's wholehearted identification with psychic elements which belong to her volitional identity *(Def. 4.1)*.

With my proposal, I mainly attempted to reply in an effective way to Watson-style objections to a hierarchical theory of freedom. According to these objections, satisfaction as the mere non-occurrence of negative attitudes toward a desire is insufficient to ground identification and free will. I replied to these objections by contending that an agent is free when she is identified with a desire *because such a desire is part of her volitional identity*. The agent's desires can ground freedom of the will not simply because, as suggested by Frankfurt, they are higher in the hierarchical structure of the will, or because the agent is satisfied with them, but because they are part of the agent's volitional identity.

Next, I continued to build my proposal by exploring the problem of ambivalence in a hierarchical theory of freedom, and I suggested a way to overcome this motivational uncertainty according to my account. Moreover, I advanced a proposal about the role of practical reason in a hierarchical theory of freedom. Finally, I challenged my account by considering cases of manipulation and addiction, and I suggested an original solution to these problems.

4

Conclusions

The following may serve as a summary of the three major claims defended in this dissertation:

I. **The problem of free will is a problem of grounding**. In this dissertation, I propose to understand free will in the terms suggested by the concept of *grounding*, and I articulate the relevant claims about freedom in such terms. The aim of this shift of perspective is to contend that when we deal with the problem of free will we are not only looking for necessary and sufficient conditions to justify it. What we are mainly interested in are elements *in virtue of which* a choice or an action can be said free. In light of this, I describe the main question about freedom as a question about which elements ground the ability to exercise free will. To ask what grounds freedom of the will, then, means to ask what circumstance or condition can make the case that the agent is exercising freedom. In other words, when we ask if an agent is choosing freely, what we aim to find is one or more conditions in virtue of which her choice or an action is a free one.

II. **Frankfurt's hierarchical theory of freedom lacks a proper grounding element.** I propose an original analysis of Frankfurt's theory of free will and its later developments. The objective of my analysis is to show that Frankfurt's classic hierarchical theory of free will is ambiguous between different definitions of freedom, and that it does not have the resources to address some very well-known objections against it. At the same time, I contend that the later developments of Frankfurt's account of freedom, that are intended as providing a reply to the objections, are not successful in that task. I explore a variety of arguments and theoretical alternatives to investigate the outcome of Frankfurt's theory. My main conclusion is that what is essentially missing in Frankfurt's theory of freedom is an effective grounding element, which can fully motivate why agents exercise free will according to a hierarchical account. As a result, I argue that Frankfurt's concept of freedom is doomed to failure, because it is left without a secure grounding element which explains the ability to exercise free will.

III. **Volitional identity is the right grounding element for free will.** I contend that an agent is free when she is identified with a desire *because* such a desire is part of her *volitional identity*, which is in turn shaped by her higher-order reflection about what she cares and loves. The agent's desires can ground freedom of the will not simply because they are higher in the hierarchical structure of the will, or because the agent is satisfied with them, but because they are part of the subject's volitional identity. For the agent to identify with some desires and not with others is, indeed, a kind of necessity, a volitional necessity. In light of this, I define free will as the agent's wholehearted identification with psychic elements which belong to her volitional identity.

The entire landscape of issues in the domains of freedom of the will, philosophy and psychology of action and motivation and related metaphysical issues is immense. Developing my original analysis and proposal in this dissertation necessarily involved ignoring, or only tangentially addressing, a

great number of important topics which would also have deserved serious attention. I hope, nonetheless, that this work might at least contribute to shedding light on some of the old problems in these domains.

Bibliography

Ayer, Alfred J. (1954). "Freedom and Necessity". In A. J. Ayer, *Philosophical Essays*. New York: St. Martin's Press, 3-20.

Beebee. Helen (2013). *Free will: an introduction.* London: Palgrave Macmillan.

Berofsky, Bernard (1983). "Autonomy". In *How Many Questions: Essays in Honor of Sidney Morgenbesser*, edited by L. Cauman, I. Levi, C. Parsons, and R. Schwartz. Indianapolis: Hackett Publishing, 301-20.

Bransen, Jan (1996). "Identification and the Idea of an Alternative of Oneself". *European Journal of Philosophy*, 4(1): 1-16.

Bratman, Michael E. (2002). "Hierarchy, Circularity and Double Reduction". In Buss & Overton (2002), 65-85.

——— (1996). "Identification, Decision and Treating as a Reason". *Philosophical Topics,* 24: 1-18.

——— (2007). *Structures of Agency*. Oxford: Oxford University Press.

Broad, Charlie D. (1934). *Determinism, Indeterminism, and Libertarianism*. Cambridge: Cambridge University Press. Reprinted in C. D. Broad, *Ethics and the History of Philosophy.* London: Routledge & Kegan Paul, 1952, 195-217.

Buss, Sarah & Overton, Lee, eds. (2002). *Contours of Agency: Essays for Harry Frankfurt*. Cambridge, Massachusetts: MIT Press.

Cuyper, Stephen (1998). "Harry Frankfurt on the Will, Autonomy and Necessity". *Ethical Perspectives*, 5(1): 44-52.

Davidson, Donald (1973). "Freedom to Act". In *Essays on Freedom and Action,* edited by T. Honderich. London: Routledge & Kegan Paul, 67–86. Reprinted in Davidson (1980), 63–81.

Davidson, Donald (1980). *Essays on Actions and Events*. Oxford: Claredon Press.

Dennett, Daniel. (1984). "I Could Not Have Done Otherwise – So What?". *The Journal of Philosophy*, 81(10): 553-65.

Dill, Brendan & Holton, Richard (2014). "The Addict in Us All". *Frontiers in Psychiatry*, 5(139): 1-20.

Dworkin, Gerald. (1970). "Acting Freely". *Nous*, 4: 367-83.

Fara, Michael (2008). "Masked Abilities and Compatibilism". *Mind*, 117: 843–65.

Fischer, John Martin (1994). *The Metaphysics of Free Will*, Oxford: Blackwell.

——— (2004). "Responsibility and Manipulation". *Journal of Ethics*. 8 (2): 145–77.

Fischer, John M. & Ravizza, Mark (1998). *Responsibility and Control: An Essay on Moral Responsibility*. Cambridge: Cambridge University Press.

Frankfurt, Harry (1969). "Alternate Possibilities and Moral Responsibility". *Journal of Philosophy*, 66(23): 829-39. Reprinted in Frankfurt (1988a), 1-10.

——— (1971). "Freedom of the Will and the Concept of a Person". *Journal of Philosophy*, 68(1): 5-10. Reprinted in Frankfurt (1988a), 11-25.

——— (1975). "Three Concepts of Free Action". *Proceedings of the Aristotelian Society*, suppl., 49: 113–25. Reprinted in Frankfurt (1988a), 47-57.

——— (1977). "Identification and Externality". In *The Identities of Persons*, edited by A. Rorty. Los Angeles: University of California Press, 239-52. Reprinted in Frankfurt (1988a), 58-68.

——— (1978). "The Problem of Action". *American Philosophical Quarterly*, 15(2): 157-62. Reprinted in Frankfurt (1988a), 69-79.

——— (1982). "The Importance of What We Care About". *Synthese*, 53(2): 257-72. Reprinted in Frankfurt (1988a), 80-94.

——— (1987). "Identification and Wholeheartedness". In *Responsibility, Character and the Emotions: New Essays in Moral Psychology*, edited by F. Schoeman. New York: Cambridge University Press, 27-45. Reprinted in Frankfurt (1988a), 159-76.

——— (1988a). The Importance of What We Care About: Philosophical Essays. Cambridge: Cambridge University Press.

——— (1988b). "Rationality and the Unthinkable". In Frankfurt (1988a), 177-90.

——— (1992). "The Faintest Passion". *Proceedings and Addresses of the American Philosophical Association*, 66(3): 5-16. Reprinted in Frankfurt (1999a), 95-107.

——— (1993). "On the necessity of ideals". In *The Moral Self*, edited by G. C. Noam, and T. Wren. Cambridge, Massachusetts: The MIT Press, 16-27. Reprinted in Frankfurt (1999a), 108-16.

——— (1999a). *Necessity, Volition, and Love*. Cambridge: Cambridge University Press.

——— (1999b). "On Caring". In Frankfurt (1999a), 155-80.

——— (1999c). "Autonomy, Necessity and Love". In Frankfurt (1999a), 129-54.

——— (2002a). "Reply to Michael E. Bratman". In Buss & Overton (2002), 86-90.

——— (2002b). "Reply to J. David Velleman". In Buss & Overton (2002), 124-28.

——— (2002c). "Reply to Gary Watson". In Buss & Overton (2002), 160-64.

——— (2002d). "Reply to T. M. Scanlon". In Buss & Overton (2002),184-88.

——— (2002e). "Reply to Richard Moran". In Buss & Overton (2002),218-25.

——— (2004). *The Reasons of Love*. Princeton: Princeton University Press.

Ginet, Carl (1990). *On Action*. Cambridge: Cambridge University Press.

Hobbes, Thomas (1999). *Treatise: Of Liberty and Necessity* [1654]. In *Hobbes and Bramhall on Liberty and Necessity*, edited by V. Chappell. Cambridge: Cambridge University Press, 15-42.

Hume, David (2007). *A Treatise of Human Nature* [1739-40]. In D. Hume, *A Treatise of Human Nature: A Critical Edition*, edited by D. F. Norton, and M. J. Norton. Oxford, Clarendon Press.

Kane, Robert H., (1996). *The Significance of Free Will*. Oxford: Oxford University Press.

——— (2002a). *The Oxford Handbook of Free Will*. Oxford: Oxford University Press.

———, ed. (2002b) "Free will, determinism, and indeterminism". In *Between Chance and Choice: Interdisciplinary Perspectives on Determinism,* edited by H. Atmanspacher, and R. C. Bishop. Thorverton: Imprint Academic, 371-406.

Lehrer, Keith (1976). "'Can' in Theory and Practice: A Possible Worlds Analysis". In *Action Theory: Proceedings of the Winnipeg Conference on Human Action*, edited by M. Brand, and D. Walton. Dordrecht: Reidel, 241-70.

Lehrer, Keith (1980). "Preferences, Conditionals, and Freedom". In *Time and Cause*, edited by P. van Inwagen. Dordrecht: Reidel, 187-220.

Levin, Michael (1979). *Metaphysics and the Mind-Body Problem*. Oxford: Oxford University Press.

Lewis, David (1981). "Are We Free to Break the Laws?". *Theoria,* 47: 113–21.

Locke, Don (1975). "Three Concepts of Free Action I". *Proceedings of the Aristotelian Society*, suppl., 49: 95-126.

Locke, John (1974). *An Essay Concerning Human Understanding* [1690]. Oxford: Claredon Press.

McKenna, Michael (2002). "Contemporary Compatibilism: Mesh Theories and Reasons-Responsive Theories". In Kane (2002), 175-98.

Moran, Richard (2002). "Frankfurt on Identification: Ambiguities of Activity in Mental Life". In Buss & Overton (2002), 189-217.

Mele, Alfred R. (1995). *Autonomous Agents*. Oxford: Oxford University Press.

——— (2003). *Motivation and Agency*. Oxford: Oxford University Press.

Moran, Richard (2002). "Frankfurt on Identification: Ambiguities of Activity in Mental Life". In Buss & Overton (2002), 189-217.

Neely, Wright (1974). "Freedom and Desire". *Philosophical Review*, 83: 32-54.

Pereboom, Derk (2001). *Living Without Free Will*. Cambridge: Cambridge University Press.

Raven, Michael J. (2015). "Ground". *Philosophy Compass*, 10(5): 322-333.

Rostbøll, Christian F. (2004). "Freedom as Satisfaction? A Critique of Frankfurt's Hierarchical Theory of Freedom". *SATS* 5(1): 131-46.

Sartorio, Carolina (2007). "Causation and Responsibility". *Philosophy Compass*, 2(5): 749–65.

——— (2012). "Causation and Freedom". *Journal of Philosophy*, 109(11): 629-51.

——— (2016a). *Causation and Free Will*. Oxford: Oxford University Press.

——— (2016b). "A Partial Defense of the Actual-Sequence Model of Freedom". *Journal of Ethics*, 20(1-3): 107-20.

Scanlon, Thomas M. (2002). "Reasons and Passions". In Buss & Overton (2002),

Segerberg, Krister (1983). "Could Have But Did Not". *Pacific Philosophical Quarterly*, 64: 230-41.

Shatz, David (1985). "Free will and the structure of motivation". *Midwest Studies in Philosophy*, 10(1): 451-82.

Slote, Michael (1980). "Understanding Free Will". *Journal of Philosophy*, 67: 136-51.

Smith, Michael (2003). "Rational Capacities, or: How to Distinguish Recklessness, Weakness, and Compulsion". In *Weakness of Will and Practical Irrationality,* edited by S. Stroud, and C. Tappolet, Christine. New York: Oxford University Press, 17–38.

Stump, Eleonore (1988). "Sanctification, Hardening of the Heart, and Frankfurt's Concept of Free Will". *Journal of Philosophy,* 85(8): 395-420.

Taylor, James, (2005). "Willing Addicts, Unwilling Addicts, and Acting of One's Own Free Will". *Philosophia*, 33(1): 237-62.

Taylor, Richard (1992). *Metaphysics*. New Jersey: Prentice Hall.

Van Inwagen, Peter (1983). *An Essay on Free Will*. Oxford: Clarendon Press.

Vellemann, David J. (1992). "What Happens when Someone Acts?". *Mind*, 101(403): 461–81. Reprinted in Vellemann (2000a), 123-43.

——— (2000a). *The Possibility of Practical Reason*. Oxford: Oxford University Press

——— (2000b). "Introduction". In Vellemann (2000a), 1-31.

——— (2002). "Identification and Identity". In Buss & Overton (2002), 91-124.

Vihvelin, Kadri (2004). "Free Will Demystified: A Dispositional Account". *Philo- sophical Topics*, 32: 427–50.

——— (2013). Causes, Laws, and Free Will: Why Determinism Doesn't Matter. New York: Oxford University Press.

Watson, Gary, (1975). "Free Agency". *Journal of Philosophy*, 72: 205–20.

——— (1987). "Free action and free will". *Mind*, 96: 154-72.

——— (2004). *Agency and Answerability*. New York: Oxford University Press.

Wolf, Susan (1990). *Freedom within Reason*. Oxford: Oxford University Press.

Zimmerman, David (1981). "Hierarchical Motivation and Freedom of the Will". *Pacific Philosophical Quarterly*, 52: 354-68.